

THE ACCOUNTING REVIEW

VOLUME 86, No. 1 JANUARY 2011

JANE THAYER

Determinants of Investors' Information Acquisition: Credibility and Confirmation
(2009 American Accounting Association Competitive Manuscript Award Winner)

JENNIFER L. BROWN

The Spread of Aggressive Corporate Tax Reporting: A Detailed Examination
of the Corporate-Owned Life Insurance Shelter

DAN S. DHALIWAL, OLIVER ZHEN LI, ALBERT TSANG, and YONG GEORGE YANG

Voluntary Nonfinancial Disclosure and the Cost of Equity Capital: The Initiation
of Corporate Social Responsibility Reporting

MICHAEL S. DRAKE, LYNN REES, and EDWARD P. SWANSON

Should Investors Follow the Prophets or the Bears? Evidence on the Use of Public
Information by Analysts and Short Sellers

MERLE EDERHOF

Incentive Compensation and Promotion-Based Incentives of Mid-Level
Managers: Evidence from a Multinational Corporation

MOHAMED Z. ELBASHIR, PHILIP A. COLLIER, and STEVE G. SUTTON

The Role of Organizational Absorptive Capacity in Strategic Use of Business
Intelligence to Support Integrated Management Control Systems

W. BROOKE ELLIOTT, JESSEN L. HOBSON, and KEVIN E. JACKSON

Disaggregating Management Forecasts to Reduce Investors' Susceptibility
to Earnings Fixation

NADER HAFZALLA, RUSSELL LUNDHOLM, and E. MATTHEW VAN WINKLE

Percent Accruals

WAYNE R. LANDSMAN, BRUCE L. MILLER, KEN PEASNELL, and SHU YEH

Do Investors Understand Really Dirty Surplus?

ALASTAIR LAWRENCE, MIGUEL MINUTTI-MEZA, and PING ZHANG

Can Big 4 versus Non-Big 4 Differences in Audit-Quality Proxies Be Attributed
to Client Characteristics?

SHU LIN, MINA PIZZINI, MARK VARGUS, and INDRANIL R. BARDHAN

The Role of the Internal Audit Function in the Disclosure of Material Weaknesses

CHRISTINE PETROVITS, CATHERINE SHAKESPEARE, and AIMEE SHIH

The Causes and Consequences of Internal Control Problems in Nonprofit
Organizations



A JOURNAL OF THE
AMERICAN ACCOUNTING ASSOCIATION

AMERICAN ACCOUNTING ASSOCIATION

The members of the 2010-2011 Executive Committee of the Association are as follows:

President	Kevin D. Stocks, Brigham Young University
President-Elect	Gregory B. Waymire, Emory University
Past President	Nancy A. Bagranoff, Old Dominion University
Vice President	Ira Solomon, University of Illinois at Urbana-Champaign
Vice President-Education	Belverd E. Needles, Jr., DePaul University
Vice President-Education-Elect	Alan Reinstein, Wayne State University
Vice President-Finance	Stacy E. Kovar, Kansas State University
Vice President-International	Kazuo Hiramatsu, Kwansei Gakuin University
Vice President-Professional Relations	Robert H. Colson, Grant Thornton, LLP
Vice President-Publications	Jean C. Bedard, Bentley University
Vice President-Publications-Elect	Christine A. Botosan, The University of Utah
Vice President-Research	John Christensen, University of Southern Denmark
Vice President-Sections and Regions	Gail Hoover King, Rockhurst University

The American Accounting Association was founded in 1916 as the American Association of University Instructors in Accounting. Its present name was adopted in 1936. The Association is a voluntary organization of persons interested in accounting education, research, and practice.

Membership in the Association is available at the following annual rates:

Full membership:

The membership dues include *Accounting Education News*, and online selection of one, two, or all three of the following Association-wide journals: *The Accounting Review*, *Accounting Horizons*, and *Issues in Accounting Education*. (This does not include Section journals.) Select from:

<i>The Accounting Review</i> only	U.S. \$185.00	with Print \$220.00
<i>Accounting Horizons</i> OR <i>Issues in Accounting Education</i> only	U.S. \$185.00	with Print \$200.00
<i>The Accounting Review</i> AND <i>Accounting Horizons</i> OR <i>Issues in Accounting Education</i>	U.S. \$195.00	with Print \$245.00
<i>Accounting Horizons</i> AND <i>Issues in Accounting Education</i>	U.S. \$195.00	with Print \$225.00
All three Association-wide journals	U.S. \$205.00	with Print \$270.00

Associate membership:

Full-time students are eligible for associate membership. The print option is not available for Associate members. Select from:

Any one Association-wide journal	U.S. \$25.00
Any two Association-wide journals	U.S. \$35.00
All three Association-wide journals	U.S. \$45.00

In addition, full members may belong to one or more Sections: Accounting, Behavior and Organizations; Accounting Programs Leadership Group; American Taxation Association; Auditing; Diversity; Financial Accounting and Reporting; Forensic and Investigative Accounting; Gender Issues and Worklife Balance; Government and Non-profit; Information Systems; International Accounting; Management Accounting; Strategic and Emerging Technologies; Teaching, Learning and Curriculum; and Two-Year College. Each Section assesses its own annual dues, which are collected by the Association.

Editorial correspondence and all manuscripts should be emailed to tar@mcombs.utexas.edu. If it is necessary to send any hard copy correspondence, the mailing address is Mary Capps, Editorial Assistant, *The Accounting Review*, The University of Texas at Austin, Department of Accounting, CBA 4M202, University Station B6400, Austin, TX 78712-0211 • Telephone: (512) 471-5318 • Fax: (512) 471-3904.

The submission fee of \$200.00 for members of the AAA or \$400.00 for nonmembers should be paid online (VISA or MasterCard only) at: <https://aaahq.org/AAAforms/journals/tarsubmit.cfm> and must be paid when article is submitted. Request for reprint permission, subscriptions, advertisements, and dues should be sent to American Accounting Association, 5717 Bessie Drive, Sarasota, FL 34233-2399 • Telephone: (941) 921-7747 • Fax: (941) 923-4093 • Email: Info@aaahq.org. Notice of change of address should be mailed at least eight weeks in advance to ensure undelayed receipt of *The Accounting Review*.

The Accounting Review is indexed in the Social Science Citation Index which is included in Web of Science.

The Accounting Review (ISSN 0001-4826 print and ISSN 1558-7967 online) is published six times a year in January, March, May, July, September, and November by the American Accounting Association, 5717 Bessie Drive, Sarasota, Florida 34233-2399. General subscriptions are available at the annual rates of \$395 for a print copy or at \$445 for a print and electronic copy. The member subscription rate for a hard copy is \$35. Electronic access is included with membership dues. Periodical postage paid at Sarasota, Florida and additional mailing offices. **POSTMASTER:** Send address change to *The Accounting Review*, 5717 Bessie Drive, Sarasota, Florida 34233-2399.

CHK-112081-56-P032760

THE ACCOUNTING REVIEW

A JOURNAL OF THE
AMERICAN ACCOUNTING ASSOCIATION

Senior Editor
Steven J. Kachelmeier

Editors

Shannon W. Anderson	D. Paul Newman
J. Harry Evans III	Thomas C. Omer
James E. Hunton	Wayne B. Thomas
Kathryn Kadous	Mark A. Trombley
Sanjay G. Kallapur	Beverly R. Walther
Ranjani Krishnan	Paul A. Zarowin
Laureen A. Maines	

56

VOL. 86

JANUARY 2011

No. 1

JANE THAYER

Determinants of Investors' Information Acquisition: Credibility and Confirmation
(2009 American Accounting Association Competitive Manuscript Award Winner) 1

JENNIFER L. BROWN

The Spread of Aggressive Corporate Tax Reporting: A Detailed Examination
of the Corporate-Owned Life Insurance Shelter 23

DAN S. DHALIWAL, OLIVER ZHEN LI, ALBERT TSANG, and YONG GEORGE YANG

Voluntary Nonfinancial Disclosure and the Cost of Equity Capital: The Initiation
of Corporate Social Responsibility Reporting 59

MICHAEL S. DRAKE, LYNN REES, and EDWARD P. SWANSON

Should Investors Follow the Prophets or the Bears? Evidence on the Use of Public
Information by Analysts and Short Sellers 101

MERLE EDERHOF

Incentive Compensation and Promotion-Based Incentives of Mid-Level
Managers: Evidence from a Multinational Corporation 131

MOHAMED Z. ELBASHIR, PHILIP A. COLLIER, and STEVE G. SUTTON

The Role of Organizational Absorptive Capacity in Strategic Use of Business
Intelligence to Support Integrated Management Control Systems 155

W. BROOKE ELLIOTT, JESSEN L. HOBSON, and KEVIN E. JACKSON

Disaggregating Management Forecasts to Reduce Investors' Susceptibility
to Earnings Fixation 185

NADER HAFZALLA, RUSSELL LUNDHOLM, and E. MATTHEW VAN WINKLE

Percent Accruals 209

WAYNE R. LANDSMAN, BRUCE L. MILLER, KEN PEASNELL, and SHU YEH

Do Investors Understand Really Dirty Surplus? 237

ALASTAIR LAWRENCE, MIGUEL MINUTTI-MEZA, and PING ZHANG

Can Big 4 versus Non-Big 4 Differences in Audit-Quality Proxies Be Attributed
to Client Characteristics? 259

SHU LIN, MINA PIZZINI, MARK VARGUS, and INDRANIL R. BARDHAN

The Role of the Internal Audit Function in the Disclosure of Material Weaknesses 287

CHRISTINE PETROVITS, CATHERINE SHAKESPEARE, and AIMEE SHIH

The Causes and Consequences of Internal Control Problems in Nonprofit
Organizations 325

BOOK REVIEWS, Stephen A. Zeff, Editor

Matthew Gill

Accountants' Truth: Knowledge and Ethics in the Financial World ANDREW ABBOTT 359

Editorial Policy and Style Information 363

Editorial Staff of *The Accounting Review*
2010–2011

SENIOR EDITOR

Steven J. Kachelmeier, The University of Texas at Austin

EDITORS

Shannon W. Anderson, University of California, Davis

J. Harry Evans III, University of Pittsburgh

James E. Hinton, Bentley University

Kathryn Kadous, Emory University

Sanjay G. Kallapur, Indian School of Business

Ranjani Krishnan, Michigan State University

Laureen A. Maines, Indiana University

D. Paul Newman, The University of Texas at Austin

Thomas C. Omer, Texas A&M University

Wayne B. Thomas, University of Oklahoma

Mark A. Trombley, The University of Arizona

Beverly R. Walther, Northwestern University

Paul A. Zarowin, New York University

EDITORIAL ASSISTANT

Mary C. Capps, The University of Texas at Austin

RESEARCH ASSISTANT

Tracie Majors, The University of Texas at Austin

American Accounting Association

EXECUTIVE DIRECTOR

Tracey E. Sutherland

COMMUNICATIONS DEPARTMENT

Beverly Harrelson

Diane Hazard

Lisa Habbnitz

Lindsay Wharton

Kathy Casper

Judy Cothorn

Sonja Langley

Jan Kovarik (subcontracted)

Copyediting Team (subcontracted)

OTHER AAA STAFF

Dierdre Harris

David Boynton

Alex Metz

Pat Stein

Debora Gardner

Peggy Turczyn

Barbara Gutierrez

Barbara Brady

Arlene Oliver

Roland LaTulip

Diane Leger

Man Fong

Jodie M. Morrill

Michelle Bowles

Kurt Gardner

Editorial Advisory and Review Board

Anwer S. Ahmed, Texas A&M University

Ashiq Ali, The University of Texas at Dallas

Anil Arya, The Ohio State University

Benjamin C. Ayers, University of Georgia

William R. Baber, Georgetown University

Stephen P. Baginski, University of Georgia

Ramji Balakrishnan, The University of Iowa

Tim Baldenius, Columbia University

Linda Smith Bamber, University of Georgia

Orie E. Barron, The Pennsylvania State University

Anne L. Beatty, The Ohio State University

Jean C. Bedard, Bentley University

Robert J. Bloomfield, Cornell University

Jennifer L. Blouin, University of Pennsylvania

Mark T. Bradshaw, Boston College

Lawrence D. Brown, Georgia State University

Brian J. Bushee, University of Pennsylvania

Jeffrey L. Callen, University of Toronto

Joseph V. Carcello, The University of Tennessee

Mary Ellen Carter, Boston College

Paul K. Chaney, Vanderbilt University

Qiang Cheng, University of Wisconsin–Madison

Michael B. Clement, The University of Texas at Austin

William M. Cready, The University of Texas at Dallas

Antonio Davila, IESE Business School–University of Navarra

Patricia M. Dechow, University of California, Berkeley

Mark L. DeFond, University of Southern California

Hemang Desai, Southern Methodist University

Sunil Dutta, University of California, Berkeley

Leslie G. Eldenburg, The University of Arizona

Patricia M. Fairfield, Georgetown University

David B. Farber, University of Missouri–Columbia

Paul E. Fischer, The Pennsylvania State University

Jennifer Francis, Duke University

Jere R. Francis, University of Missouri–Columbia

James R. Frederickson, Melbourne Business School

Frank B. Gigler, University of Minnesota

Cristi A. Gleason, The University of Iowa

Johnathan C. Glover, Carnegie Mellon University

Steven M. Glover, Brigham Young University

David A. Guenther, University of Oregon

Luzi Hail, University of Pennsylvania

Jeffrey Hales, Georgia Institute of Technology

John R. M. Hand, The University of North Carolina at Chapel Hill

R. Lynn Hannan, Georgia State University

Gilles Hilary, INSEAD

D. Eric Hirst, The University of Texas at Austin

Frank D. Hodge, University of Washington

Chris E. Hogan, Michigan State University

Ole-Kristian Hope, University of Toronto

Richard W. Houston, The University of Alabama

Raffi J. Indjejikian, University of Michigan

Christopher D. Ittner, University of Pennsylvania

Debra C. Jeter, Vanderbilt University

Karla M. Johnstone, University of Wisconsin–Madison

Bin Ke, Nanyang Technological University

S. Jane Kennedy, University of Washington

Inder K. Khurana, University of Missouri–Columbia

Michael D. Kimbrough, University of Maryland

Ronald R. King, Washington University in St. Louis
April Klein, New York University
W. Robert Knechel, University of Florida
Lisa L. Koonce, The University of Texas at Austin
Susan D. Krische, American University
Jayanthi Krishnan, Temple University
William N. Lanen, University of Michigan
David F. Larcker, Stanford University
Reuven Lehavy, University of Michigan
Carolyn B. Levine, Carnegie Mellon University
Robert Libby, Cornell University
Joan L. Luft, Michigan State University
Anne M. Magro, George Mason University
Carol Marquardt, Baruch College-CUNY
Dawn A. Matsumoto, University of Washington
Edward Maydew, The University of North Carolina at Chapel Hill
Brian W. Mayhew, University of Wisconsin-Madison
Sarah E. McVay, University of Utah
Kenneth A. Merchant, University of Southern California
Gregory S. Miller, University of Michigan
Lillian F. Mills, The University of Texas at Austin
Brian Mittendorf, The Ohio State University
Partha S. Mohanram, University of Toronto
Steven J. Monahan, INSEAD
Donald V. Moser, University of Pittsburgh
James N. Myers, University of Arkansas
Venky Nagar, University of Michigan
Terry L. Neal, The University of Tennessee
Karen K. Nelson, Rice University
Edward F. O'Donnell, Southern Illinois University Carbondale
Per M. Olsson, Duke University
Chul W. Park, The University of Hong Kong
Mark E. Peecher, University of Illinois at Urbana-Champaign
Sean A. Pfeffer, University of Kentucky

Kathy Ruby Petroni, Michigan State University
John D. Phillips, University of Connecticut
Morton Pincus, University of California, Irvine
Joseph D. Piotroski, Stanford University
Marlene A. Plumlee, University of Utah
Suresh Radhakrishnan, The University of Texas at Dallas
Kannan Raghunandan, Florida International University
Shiva Rajgopal, Emory University
Scott Richardson, BlackRock Inc.
John R. Robinson, The University of Texas at Austin
Darren T. Roulstone, The Ohio State University
Katherine Schipper, Duke University
Karen L. Sedatole, Michigan State University
Catherine M. Shakespeare, University of Michigan
Lakshmanan Shivakumar, London Business School
K. (Shiva) Sivaramakrishnan, University of Houston
Hollis A. Skaife, University of Wisconsin-Madison
Richard G. Sloan, University of California, Berkeley
J. Reed Smith, Indiana University at Indianapolis
Mark T. Soliman, University of Washington
Geoffrey B. Sprinkle, Indiana University
Mary Harris Stanford, Texas Christian University
Phillip C. Stocken, Dartmouth College
K. R. Subramanyam, University of Southern California
Hun-Tong Tan, Nanyang Technological University
Kristy L. Towry, Emory University
Senyo Tse, Texas A&M University
Jennifer Wu Tucker, University of Florida
Irem Tuna, London Business School
Mohan Venkatachalam, Duke University
Joseph P. Weber, Massachusetts Institute of Technology
Michael S. Wilkins, Texas A&M University
Richard H. Willis, Vanderbilt University
Arnold M. Wright, Northeastern University
Teri Lombardi Yohn, Indiana University
Xiao-Jun Zhang, University of California, Berkeley
Mark F. Zimbelman, Brigham Young University

PAST EDITORS

1926–1929	William A. Paton	1976–1978	Don T. DeCoster
1929–1943	Eric L. Kohler	1978–1983	Stephen A. Zeff
1944–1947	A. C. Littleton	1983–1987	Gary L. Sundem
1948–1949	Robert L. Dixon	1987–1990	William Kinney
1950–1959	Frank P. Smith	1990–1994	A. Rashad Abdel-khalik
1959–1962	Robert K. Mautz	1994–1997	Robert P. Magee
1962–1965	Lawrence L. Vance	1997–2000	Gerald L. Salamon
1965–1967	Wendell Trumbull	2000–2002	Linda Smith Bamber
1968–1970	Charles H. Griffin	2002–2005	Terry Shevlin
1971–1972	Eldon S. Hendriksen	2005–2008	Dan S. Dhaliwal
1973–1975	Thomas F. Keller		

2009 American Accounting Association
Competitive Manuscript Award Winner

Determinants of Investors' Information Acquisition: Credibility and Confirmation

Jane Thayer
The University of Georgia

ABSTRACT: Psychology theory suggests that current investors are unlikely to be impartial in their acquisition of information regarding a currently held position after receiving information that casts doubt on the profitability of the position. This study examines whether the favorability of initial information an investor receives subsequent to taking an investment position moderates his acquisition of additional information regarding that position. A web-based experiment indicates that participants generally choose to view information based on its credibility. However, the majority of information chosen by participants initially receiving unfavorable information supports the investment position they chose at the beginning of the case. Moreover, participants receiving unfavorable information prefer to view lower-credibility, preference-consistent information at a similar rate to higher-credibility, preference-inconsistent information. Given a psychological need to support a previous investment decision, participants forgo a certain amount of credibility in the information they gather to confirm their belief in their investment position.

Keywords: *investors; information acquisition; preferences; credibility; experiment.*

Data Availability: *Contact the author.*

I. INTRODUCTION

Investors have an economic incentive to be objective and impartial in their acquisition of information, as a high-quality set of information is central to an optimal investment decision. Psychology research suggests, however, that impartiality will be more difficult after an investment position has been taken (e.g., Festinger 1957; Leventhal and Brehm 1962; Festinger 1964).

This paper is based on my dissertation completed at Emory University. I thank my dissertation chairperson, Kathryn Kadous, for her support and guidance throughout this project, and my other dissertation committee members, Kristy Towry and Greg Waymire, for their many helpful suggestions. This paper also benefited from the valuable comments of Ben Ayers, Ann Backof, Liada Bamber, Michael Bamber, Brooke Elliott, Sukari Farrington, Jeffrey Hales, Jackie Hammersley, Gary Hecht, Max Hewitt, Pat Hopkins, Lisa Koonce, Susan Krische, Bob Libby, Laureen Maines, Molly Mercer, Mark Nelson, Derek Oler, Mark Peecher, Grace Pownall, Bill Tayler, Xue Wang, Donnie Young, and workshop participants at Carnegie Mellon University, Cornell University, Emory University, Georgia Institute of Technology, Indiana University, The University of Georgia, University of Illinois at Urbana-Champaign, University of Massachusetts Amherst, The University of Texas at Austin, and the 2008 American Accounting Association Annual Meeting. I am grateful for the financial support of the Roberto C. Goizueta Foundation and the Professor Jagdish and Mrs. Madhuri Sheth Fund.

Editor's note: Accepted by Laureen Maines.

Submitted: February 2009

Accepted: June 2010

Published Online: January 2011

An underlying assumption of cognitive dissonance theory is that individuals prefer their actions to match their beliefs (Festinger 1957). Information received subsequent to taking a position influences whether this internal balance is maintained (Festinger 1957, 1964). If information casts doubt on the future profitability of an investor's position, then psychology theory suggests that the investor may selectively seek additional information supporting the position to regain that balance (e.g., Frey 1986). This study examines whether the initial information an investor receives subsequent to taking an investment position influences the type of information s/he acquires to make future judgments and decisions about the position. Will an investor acquire a balanced set of credible information or will s/he acquire information that confirms his/her investment position?¹

Investors' reactions to information based on its source credibility (e.g., Clement and Tse 2003; Gleason and Lee 2003) suggest that investors recognize the need to be accurate in their estimation of security prices. For example, investors react more quickly to forecast revisions made by celebrity analysts (Gleason and Lee 2003) and analysts employed by prestigious brokerage firms (Clement and Tse 2003) than to those made by lesser-known analysts.² In addition, investors react to the credibility of management forecasts in addition to the news, or surprise, in those forecasts (Jennings 1987; Mercer 2004; Hutton and Stocken 2009). Investors have also been shown to respond more strongly to earnings released by companies employing large, international audit firms compared to those of companies employing smaller audit firms (Teoh and Wong 1993). It is unknown, however, whether an investor's sensitivity to the credibility of information will continue after s/he receives initial information that casts doubt on a recently chosen position.

Although it is more difficult to remain objective after taking a position, research suggests that individuals are mindful of the credibility of information regarding the position, especially when the information is unfavorable (Ditto et al. 1998). However, given the choice between high-credibility information that is consistent with their preference to believe in the position (hereafter, preference-consistent information) and that which is inconsistent with their preferred belief (hereafter, preference-inconsistent information), studies show that an individual's need to support a desired belief will lead him/her to choose the former (Frey 1981, 1986).³ These findings, however, result from a context that does not include an immediate economic incentive to gather a balanced set of credible information as an aid to decision making (Frey 1981).

With regard to individuals' willingness to trade off credibility for preference-consistency, studies suggest that individuals selectively seeking preference-consistent information will prefer to receive preference-inconsistent information if it has more credibility than available preference-consistent information (Lowin 1969). However, the psychology studies investigating individuals' preferences for information provide participants with information varied between two extreme levels of credibility (e.g., Lowin 1967, 1969; Frey 1981; Ditto et al. 1998). Given the vast amount of information from various sources in the capital markets, it is unknown whether an investor would seek information based on its credibility or whether s/he would choose to acquire information supportive of his/her investment position.

¹ Regarding information acquisition, I use the term "balanced" to describe a representative sample of the information available from all sources, with no one viewpoint under- or over-represented in the sample, relative to its representation in the population of available information.

² Studies show that rankings and employer size are related to analysts' forecast accuracy (Stickel 1992; Clement 1999; Hong and Kubik 2003); therefore, these signals can provide investors with a cognitive shortcut in determining the credibility of a particular forecast.

³ The psychology literature uses the terms "preference-consistent" and "preference-inconsistent" to describe information that is congruent or incongruent, respectively, with an individual's preferred belief, regardless of when it is received in a decision-making process (e.g., Ditto and Lopez 1992; Ditto et al. 1998; Ditto et al. 2003). However, for ease of exposition, I will use the terms "favorable" or "unfavorable" to describe the valence of the *initial information* an investor receives subsequent to taking an investment position and "preference-consistent" or "preference-inconsistent" to describe the valence of the *additional information* the investor seeks subsequent to receiving initial information.

The capital markets information environment is unique, in that investors often know—before investing in detailed analysis of a financial report or news article—whether information from a given source supports a particular investment position. For example, websites such as Yahoo! Finance and Dow Jones' Market Watch offer listings of analysts' broad recommendations (e.g., buy, hold, sell) for a given stock. Dow Jones' Market Watch separates analysts' recommendations by upgrades, downgrades, and initiations on its site, providing the name of the brokerage firm, the new recommendation, the old recommendation, and a brief set of comments summarizing the analyst's report. Therefore, investors can quickly determine the flavor of these reports and may, in certain cases, select information based on its level of support for a particular investment position. Moreover, investors' ability to "pre-screen" information is not only limited to analyst reports, but is also inherent in searches for news and popular press articles through databases such as Factiva, ProQuest, and Lexis-Nexis. Federated search engines for business information, such as www.Biznar.com, also allow individuals to search a topic and receive an article title and source.

Given the psychological need to find support for a belief in a prior action when the action is called into question (e.g., Festinger 1957, 1964; Frey 1986) and the nature of today's information environment, I predict that investors seeking support for a previous investment decision after the receipt of unfavorable information will acquire additional information that confirms their position. Specifically, I predict that these investors will choose high-credibility, preference-consistent information over high-credibility, preference-inconsistent information. Moreover, I predict that these individuals, who are in search of information to bolster their belief in their investment position, will forgo a certain amount of credibility to acquire support for that position. On the other hand, I predict that individuals who receive initial favorable information regarding their investment position will experience less of a need to support their position and will seek a balanced set of information from high-credibility sources.

I use a web-based experiment to test my predictions about investors' information acquisition after they receive initial information regarding a previously chosen investment position. Holding constant the initial information given to participants regarding their chosen investment, I manipulate the favorability of that information by manipulating two variables between participants: (1) the participant's directional investment position (long versus short) and (2) the participant's earnings benchmark (high versus low). The randomly assigned investment position (long versus short) introduces a directional preference regarding the firm's earnings outcome, and the randomly assigned earnings benchmark (high versus low) introduces a preference for a particular level of earnings. Those assigned a long (short) position are asked to choose a long (short) investment in one of two firms.

After making their investment choice and learning their earnings benchmark, participants receive initial information about their investment and are provided an opportunity to view additional information in the form of four analyst reports before making an EPS forecast. Two reports are issued by analysts employed by well-known brokerage firms (high-credibility sources) and two by unknown analysts (low-credibility sources). Within each set, one estimates high earnings and the other estimates low earnings. Participants can view as many or as few of the reports as they wish. The primary dependent measure is the time spent viewing each analyst report.

Results generally support my predictions, as I find that the information sought by participants is a function of the favorability of the initial information they receive subsequent to taking an investment position. Holding constant the amount of preference-consistent information offered by high- and low-credibility sources, participants prefer to view high-credibility information. However, participants receiving initial favorable information are more balanced in their acquisition of information based on its preference-consistency than are participants receiving unfavorable information. The majority of high-credibility information chosen by participants receiving unfavorable information is that which is preference-consistent. Additionally, participants receiving initial un-

favorable information prefer to view, at a similar rate, less credible information that is preference-consistent and more credible information that is preference-inconsistent. Consequently, participants' earnings expectations are biased in the direction of the information they choose to view.

These findings contribute to both the accounting and psychology literatures regarding the influence of individuals' preferences on information acquisition and evaluation in three primary ways. First, my study presents a scenario in which the initial information an investor receives regarding a recent investment sets the course of his/her information-acquisition process. Results suggest that an investor receiving unfavorable information regarding his/her investment position will gather additional information in such a way that the resulting information set overly represents information supporting his/her position. Additionally, some of the over-represented information may have less credibility than information that is under-represented. The outcome of this information-acquisition process is a set of information that can result in biased estimates and forgone profits.

Second, the finding that investors are mindful of a source's viewpoint when gathering information has implications for archival researchers investigating market-pricing efficiency (e.g., Diamond and Verrecchia 1987; Figlewski and Webb 1993; Clement and Tse 2003; Gleason and Lee 2003). The results of my study suggest that, under certain circumstances, investors are likely to consider an information source's viewpoint in addition to its credibility in reacting to the information (e.g., Clement and Tse 2003; Gleason and Lee 2003; Mercer 2004; Hutton and Stocken 2009). Additionally, the results present a potential reason for findings that the market's pricing of firms with short-selling constraints, which by definition lack bearish traders, is less efficient than that for firms with short-sale opportunities (Diamond and Verrecchia 1987; Figlewski and Webb 1993).

Third, the study suggests that prior findings in psychology, which find that individuals receiving preference-inconsistent information are sensitive to the credibility of that information and any additional information they might receive, represent a boundary condition (e.g., Lowin 1969; Frey 1981; Ditto et al. 1998). The information offered to participants in these psychology studies is varied between two extreme levels of credibility, resulting in participants choosing high-credibility, preference-inconsistent information over low-credibility, preference-consistent information (Frey 1981). On the other hand, the capital markets information environment offers investors information from various sources representing a wide range of credibility. My study indicates that investors who are seeking support for a desired belief and who are presented with a choice of information from various sources are likely to accept some reduction in information credibility in order to gather support for their position.

The remainder of the paper is organized as follows. Section II develops the theory and hypotheses. The experimental design and experimental results are presented in Sections III and IV, respectively. Section V summarizes and concludes.

II. THEORY AND HYPOTHESES

Source credibility is one way in which an investor's level of uncertainty in information can be reduced (Erdem and Swait 1998). This reasoning is supported by empirical evidence that investors react to information based on attributes of its source (e.g., Jennings 1987; Teoh and Wong 1993; Clement and Tse 2003; Gleason and Lee 2003; Mercer 2004; Rogers and Stocken 2005; Hutton and Stocken 2009), which suggests that investors recognize the importance of forming an accurate estimation of a security's market price. These studies, however, do not consider the current position of an investor. Psychology theory suggests that current investors are likely to find it difficult to be impartial and objective in their search for information regarding a currently held position after receiving information that casts doubt on the profitability of the position (e.g., Festinger

1957, 1964; Frey 1986). The current study examines whether the favorability of initial information an investor receives subsequent to taking an investment position moderates his/her acquisition of additional information regarding that position.

Extant research in accounting and psychology does not provide a clear prediction of investors' information-acquisition process in the course of making judgments and decisions regarding a currently held position. Hales (2007) finds that experimental participants, in the role of investors, interpret analyst forecast information based on their investment position. Specifically, Hales (2007) presents results suggesting that an investor's earnings expectation for an investee firm is more likely to agree with analyst forecast information if that information indicates a high likelihood of a potential gain as compared to that of a loss. However, participants in Hales' (2007) study were not provided with identifying information regarding the reporting analysts' credibility. Research in psychology suggests that an investor receiving unfavorable information regarding his/her investment position will be particularly sensitive to the credibility of the information (Ditto et al. 1998). In turn, s/he would adjust his/her beliefs if the unfavorable information was of high-credibility compared to low-credibility, favorable information.

On the other hand, the psychology research investigating individuals' sensitivity to information based on its credibility and preference-consistency manipulates credibility in an extreme, dichotomous manner (e.g., Lowin 1969; Frey 1981; Ditto et al. 1998). It is unclear how individuals will choose to receive information provided by different sources when the credibility of those sources is varied at more intermediate levels (Lowin 1967). My study seeks a nuanced description of individuals' information-acquisition process when there are preferences to support a desired belief and a more moderate amount of disparity in the various information sources' credibility.

Credibility and Preference Consistency

Cognitive dissonance forms the basis for the effect of directional preferences on individuals' information acquisition (Festinger 1957, 1964). Festinger (1957) describes dissonance as arising when two incompatible sets of knowledge, or cognitions, are held simultaneously. The need to reduce dissonance can lead individuals to be selective in their acquisition of additional information. Because of the opportunity for dissonance to arise upon the initial receipt of information subsequent to taking an investment position, I argue that the favorability of this information regarding the investor's position sets the course of his/her information-acquisition process.

The receipt of information supporting the likely profitability of a recently chosen investment position will corroborate the investor's decision. As such, s/he should experience less need for supportive information, instead seeking a balanced set of information from high-credibility sources regardless of the source's support of the investment position. On the other hand, the receipt of unfavorable information subsequent to an investment choice should arouse dissonance and the investor will, consequently, seek a reduction of the dissonance. One way that individuals can reduce dissonance is by searching for preference-consistent information (Brehm and Cohen 1962).

Research in psychology, however, suggests that investors selectively seeking preference-consistent information will prefer to receive preference-inconsistent information if it is of higher quality than available preference-consistent information (Frey 1981). This finding, along with research suggesting that individuals receiving high-credibility, preference-inconsistent information will update their preferred belief if additional preference-consistent information is low in credibility (Ditto et al. 1998), suggests that individuals seeking additional preference-consistent information are sensitive to source credibility. Therefore, holding constant the amount of preference-consistent information available from high-credibility and low-credibility sources, I expect investors to acquire more high-credibility information than low-credibility information. This prediction is summarized below:

- H1:** Investors receiving initial favorable or unfavorable information regarding their position will acquire more information from high-credibility sources than from low-credibility sources.

Findings in psychology suggest, however, that given a choice of information from two similarly credible sources, investors will prefer preference-consistent information to preference-inconsistent information (Lowin 1967, 1969; Frey 1986). Such a finding would indicate that an investor is collecting an unbalanced set of information, which could result in the potential over-weighting of information supportive of his/her current position in future decision making. Additionally, this result of selective acquisition would have implications for the generality of previous archival findings of investors' reaction to information based on source credibility (e.g., Clement and Tse 2003; Gleason and Lee 2003), as these studies do not control for the current position of investors.

Regardless of an investor's economic incentives, I predict that the need to support a previous investment decision will influence his/her information-acquisition process. I expect that an investor receiving initial favorable information regarding his/her investment position will seek a balanced set of both preference-consistent and preference-inconsistent information from high-credibility sources. On the other hand, I expect an investor for whom initial information is unfavorable regarding his/her position to seek high-credibility information that is preference-consistent over that which is preference-inconsistent. These predictions are summarized below:

- H2a:** Investors receiving initial favorable information regarding their position will acquire an equal set of preference-consistent and preference-inconsistent information from high-credibility sources.

- H2b:** Investors receiving initial unfavorable information regarding their position will acquire more preference-consistent information than preference-inconsistent information from high-credibility sources.

Credibility versus Preference Consistency

Previous research examining individuals' selective exposure to information shows that individuals tend to select high-credibility, preference-inconsistent information over low-credibility, preference-consistent information (Frey 1981). This is consistent with findings that information usefulness can influence information search over an individual's need to reduce dissonance (Wicklund and Brehm 1976). The fear of invalidity or a judgmental mistake can lead individuals to seek additional information they view as useful before making a conclusion (Kruglanski and Webster 1996).

The information presented to experimental participants in the psychology studies of selective exposure to information based on source credibility is varied between two extreme levels of credibility (e.g., Lowin 1969; Frey 1981). For example, both Lowin (1969) and Frey (1981) present to experimental participants information that is designed to be either maximally easy or hard to refute. These studies manipulate the messages' ease of refutation by varying the presumed expertise of the information source. Participants were told that the messages were provided by either an expert economist (i.e., low ease of refutation) or by high school sophomores (i.e., high ease of refutation).

The information environment manufactured in the psychology studies described above is very different from the capital markets information environment. With the extensive amount of information available to investors in the marketplace, the sources of information are also vast. Given the availability of information from various sources representing a wide range of credibility, an investor may decide to trade off some amount of credibility in order to gather information sup-

portive of his/her investment position, especially after the receipt of information that casts doubt on the future profitability of his/her position. I predict that investors receiving initial information that is unfavorable regarding their position will not acquire preference-inconsistent information from a high-credibility source at a greater rate than preference-consistent information from a low-credibility source. This prediction is summarized below:

H3: Investors receiving initial unfavorable information regarding their position will *not* acquire more high-credibility, preference-inconsistent information than low-credibility, preference-consistent information.

III. EXPERIMENT

Participants

To test my hypotheses, I conduct a web-based experiment with 92 second-year students from a *BusinessWeek* top-25 M.B.A. program taking on the role of investors. Approximately 58 percent of the participants report that they have made investments in individual firm stocks and 99 percent indicate that they plan to do so in the future. Those who have invested in individual stocks report having purchased a mean (median) of 55 (20) stocks. Additionally, 79 percent of participants report having previously invested in stock mutual funds.

Experimental Procedures

I designed a web-based experimental platform to test whether an investor's need to support an investment decision influences the type of information s/he acquires before completing a basic earnings-forecasting task. Each participant chooses one of two firms in which to invest. A choice is included in the task in order that the decision instills in participants a commitment to their investment position (Brehm and Cohen 1962). In making an investment choice, participants are expected to feel responsible for the future consequences of their decision.

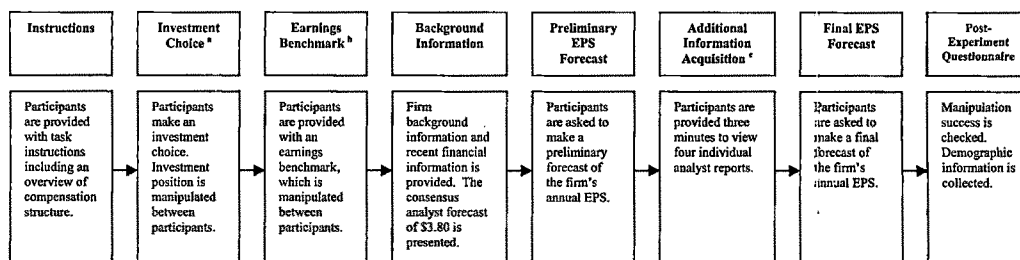
Participants are randomly assigned to either a long position or a short position before choosing a firm in which to invest. Participants view information regarding each of the firms in order to make their choice. The information is not only designed to present the firms as equally attractive, but is also ambiguous regarding the future profitability of the firms.⁴ Information presented to participants is held constant across all conditions.⁵ After making their choice, participants are randomly assigned either a high or low earnings-per-share benchmark to which actual EPS are compared in determining if they gain or lose from their individual investment.

Participants' choice of Firm X or Firm Y has no effect on the information presented subsequently. That is, all participants see the same information after their investment choice, with the exception of the firm's name. Participants receive quarterly earnings information for the past three years, recent historical stock prices over the same period, and the consensus analyst forecast for the upcoming fiscal year, including the range of analysts' individual forecasts. The consensus analyst forecast is either favorable or unfavorable, based on the individual participant's assigned investment position and earnings benchmark. After receiving this information, participants have an opportunity to view any or all of four analysts' reports before making their final EPS estimate. Figure 1 presents a timeline of the experimental task. See the Appendix for an example of the wording of the experimental manipulations.

⁴ Brehm (1956) finds that post-choice dissonance increases as the attractiveness of the rejected alternative increases. An individual choosing between two attractive alternatives has a choice to make; however, an individual selecting between one attractive alternative and one unattractive alternative will likely feel as if s/he does not have a choice.

⁵ This information, including earnings information, is modeled after that of an actual publicly traded firm.

FIGURE 1
Timeline of Experimental Task



- ^a Experiment participants assigned to the long (short) investment position condition were asked to choose a long (short) investment in one of two firms. Prior to making their investment choice, all participants viewed the same information regarding the two firms.
- ^b Participants were randomly assigned an earnings benchmark of either \$3.76 or \$3.84. Those assigned to the long investment position condition were told that they would receive a higher (lower) payoff when actual earnings were greater (less) than their earnings benchmark. Participants assigned to the short investment position condition were told that they would receive a higher (lower) payoff when actual earnings were less (greater) than their earnings benchmark.
- ^c The amount of time a participant spends viewing each of the four analyst reports is the main dependent measure.

Experimental Treatments

I employ a 2×2 full-factorial, between-participants design that manipulates (1) the participant's investment position (long versus short) and (2) their earnings benchmark (low versus high).⁶ As in Hales (2007), each participant's payoff is either a positive or a negative function of the subject firm's earnings, depending on his/her randomly assigned investment position (long versus short). After participants make their investment choice, they are randomly assigned an earnings-per-share benchmark of either \$3.76 (low) or \$3.84 (high). Those in the long investment position condition experience a higher (lower) payoff when actual earnings are greater (less) than their earnings benchmark. Participants in the short investment position condition experience a higher (lower) payoff when actual earnings are less (greater) than their earnings benchmark. The combination of participants' randomly assigned investment position and earnings benchmark determines whether the initial information received subsequent to their investment choice (i.e., the consensus analyst forecast of \$3.80) is favorable or unfavorable.⁷

After receiving the consensus analyst forecast, participants receive four analysts' forecasts and a link to view each analyst's corresponding report. Two of the individual forecasts, one from a well-known equity research firm and one from an unknown equity research firm, estimate annual

⁶ The manipulation of the two independent variables is based on Hales (2007). I am grateful to Jeffrey Hales for his generosity in allowing me access to his experimental instrument.

⁷ The consensus analyst forecast of \$3.80 foretells that participants assigned to the long (short) investment position condition and an earnings benchmark of \$3.76 (\$3.84) will experience a gain with regard to their investment position. However, this consensus analyst forecast foretells that participants assigned to the long (short) investment position condition and an earnings benchmark of \$3.84 (\$3.76) will experience a loss based on their investment position.

earnings to be at the top of the range of all analysts' individual forecasts (i.e., \$3.88). The other two forecasts, one from a well-known firm and one from an unknown firm, estimate annual earnings at the bottom of the range of forecasts (i.e., \$3.72). Results from a separate experiment indicate that the well-known firms are expected to provide more credible research than the unknown firms.⁸

The four forecasts are presented in a 2×2 matrix rather than a list in order to prevent participants from choosing the analyst reports sequentially. In addition, the computerized program randomizes, across participants, the placement of the forecasts in the matrix. In order to view a particular report, a participant must click on the link associated with that report. Participants are given up to three minutes to view any or all of the four analyst reports; however, participants who spend less than the allotted time are awarded bonus points. Therefore, at any point during these three minutes, participants have the opportunity to stop viewing this information and move forward in the task where they are asked to provide their final EPS forecast.

Dependent Measures

After receiving initial information (i.e., the consensus analyst forecast) that is either favorable or unfavorable regarding their investment position, participants have an opportunity to view four individual analyst reports. I measure the amount of time a participant spends viewing each of the four reports. This allows me to examine whether participants are choosing to view information based on its credibility or its preference consistency. An analyst report will be preference consistent, or will bolster the participant's preferred belief that s/he made a wise investment choice, if it suggests that s/he will experience a gain from the investment (i.e., more optimistic forecasts if a long position is held; less optimistic forecasts if a short position is held). The analyst report will be preference-inconsistent if it casts doubt on the future profitability of the participant's chosen investment position (i.e., less optimistic forecasts if a long position is held; more optimistic forecasts if a short position is held).

Participants' individual viewing times for each report allow me to examine their relative preferences to acquire the information. As the amount of preference-consistent information provided by high- and low-credibility sources is held constant, H1 predicts that all participants will spend more time viewing high-credibility information than low-credibility information. Hypothesis 2a predicts that participants for whom the consensus analyst forecast is favorable will spend equal time viewing the high-credibility, preference-consistent and high-credibility, preference-inconsistent analyst reports. Hypothesis 2b predicts that participants for whom the consensus analyst forecast is unfavorable will spend more time viewing the high-credibility, preference-consistent analyst report than the high-credibility, preference-inconsistent analyst report. Last, H3 predicts that participants for whom the consensus analyst forecast is unfavorable will *not* spend more time viewing the high-credibility, preference-inconsistent analyst report than the low-credibility, preference-consistent analyst report.

⁸ In a separate experiment, 41 participants from the same population as that examined in the main experiment rated the credibility of equity research they would expect to receive from each of these four firms. A five-point Likert scale was used with 1 being "not at all credible" and 5 being "highly credible." Mean ratings for the two well-known firms were 4.20 and 4.27. The difference in these ratings is not significant ($t_{40}=0.83$, $p = 0.41$). The two unknown firms received mean ratings of 3.22 and 3.27. The difference in these ratings is not significant ($t_{40}=0.36$, $p = 0.72$). However, the average credibility rating for the two well-known firms (mean = 4.23) was significantly greater than that for the unknown firms (mean = 3.24) ($t_{40}=4.60$, $p < 0.001$). Going forward, I use the labels "low-" and "high-" credibility to connote that the rating of the credibility of the research issued by the two unknown firms was significantly less than the rating of the credibility of the research issued by the two well-known firms. As such, these labels are not meant to convey absolute levels of credibility but, instead, relative levels of credibility.

Compensation

Participants receive compensation based on the number of points earned from four sources. They receive 25 participation points at the beginning of the case. Additionally, they earn points based on the accuracy of their final EPS forecast. If actual earnings equal the participant's forecasted earnings, then s/he earns 25 points; however, one point of the 25 "accuracy points" is deducted for every penny that his/her forecast differs from actual earnings. This component of the compensation scheme provides participants an incentive to make an accurate forecast.

Additionally, participants earn or lose up to 25 points based on the outcome of their investment position. Those holding a long position earn (lose) two points for every penny that actual earnings-per-share exceeds (falls below) their earnings benchmark. Similarly, participants holding a short position earn (lose) two points for every penny that actual earnings-per-share falls below (exceeds) their earnings benchmark. All gains or losses from the participant's investment position are capped at 25 points. This component of the compensation scheme captures investors' incentive to hold a profitable investment.⁹

Lastly, participants have an opportunity to receive bonus points if they do not use the entire three minutes allotted to view the analyst reports. A three-minute timer is provided at the top of the computer screen in order for participants to keep track of their time. The timer clicks down to zero before automatically advancing to the next section of the study where participants are asked to make their final EPS forecast. If participants choose to move forward with 60 seconds or more on the timer, they are awarded ten bonus points. Participants leaving between 30 and 59 seconds on the timer receive five bonus points. These points are added to the participant's total points at the end of the study. This component of the compensation scheme captures the effort expended in information acquisition.¹⁰

In sum, participants have an opportunity to earn up to 85 points. At the end of the case, each participant's points are converted into U.S. dollars. Participants are guaranteed payment of \$5, but could earn up to \$25. On average, participants receive compensation equal to \$16.¹¹

IV. RESULTS

Manipulation Checks

To verify that participants attended to the manipulation of investment position (i.e., long versus short), I asked them in a post-experiment questionnaire to indicate whether they gained or lost points if the firm in which they were invested experienced an increase in earnings.¹² This manipulation was successful: 87 percent (41/47) of participants assigned a long investment position reported that they gained points if earnings rose, while only 7 percent (3/45) of participants

⁹ To ensure that participants understand the accuracy incentive and profitability incentive components of the compensation scheme, they are required to answer correctly five questions before moving forward in the task.

¹⁰ Studies of selective exposure to information generally force participants to choose between sets of supportive and non-supportive information (e.g., Lowin 1967; Frey and Wicklund 1978; Frey and Stahlberg 1986). Instead of incorporating a similar choice into my study, I implement the time cost. The costliness of analyzing all information should induce choice. Additionally, time spent in an information-processing task is highly correlated with characteristics of effort (Bettman et al. 1990; Sprinkle 2000). Therefore, the time participants spend viewing the analyst reports is a measure of the effort they are willing to expend in gathering additional information.

¹¹ Compensation was paid at a rate of approximately \$0.29 per point earned, with a minimum compensation equal to \$5.00. Participants earned an average of 54 points.

¹² Prior to making their investment choice, all participants viewed the same information regarding the two firms. Participants' investment position condition did not influence their investment choice ($\chi^2(1) = 0.69$, $p = 0.41$). In addition, participants' compensation, based in part on their EPS estimate, was not affected by their investment choice ($t_{90} = 0.08$, $p = 0.94$). Participants choosing Firm X earned an average of 53.6 points and received approximately \$16 in compensation, while participants choosing Firm Y earned an average of 53.7 points and also received approximately \$16 in compensation.

assigned a short investment position used this response. The difference is significant (Fisher's Exact Test; $p < 0.001$). To verify that participants attended to the difference between the consensus analyst forecast (i.e., \$3.80) and their earnings benchmark (i.e., \$3.76 versus \$3.84), I asked them to provide both amounts. Eighty-nine percent (42/47) of those with an earnings benchmark of \$3.76 provided a response that was within \$0.02 of their benchmark. Ninety-six percent (43/45) of those with an earnings benchmark of \$3.84 provided a response that was within \$0.02 of their benchmark. Approximately 97 percent (89/92) of all participants were within \$0.02 of the correct consensus analyst forecast. Most importantly, 94 percent (86/92) of all participants attended to the direction of the consensus analyst forecast in relation to their earnings benchmark. Participants' awareness of the consensus analyst forecast being higher or lower than their earnings benchmark and their responses indicating their understanding that they either gained or lost points when actual earnings increased indicates that the manipulation of the favorability of the consensus analyst forecast was successful.¹³

Mixed-Design ANOVA

As the objective of the study is to examine the effect of the favorability of initial information received subsequent to an investment decision (i.e., the consensus analyst forecast of \$3.80) on investors' viewing time of additional information, I collapse the four experimental conditions into two conditions based on favorability and report the results accordingly. The Favorable condition is comprised of participants assigned to the long (short) investment position with an earnings benchmark of \$3.76 (\$3.84). The Unfavorable condition is comprised of participants assigned to the long (short) investment position with an earnings benchmark of \$3.84 (\$3.76).¹⁴ I then analyze the information using a mixed-design analysis of variance with Favorability as a between-participants variable and the Credibility (high credibility versus low credibility) and Preference Consistency (preference-consistent versus preference-inconsistent) of the analyst reports as two within-participants variables.

Main Effects of Favorability, Credibility, and Preference Consistency on Viewing Time

After receiving initial information regarding their chosen investment positions, participants had the opportunity to view four analyst reports. Based on the participant's assigned experimental condition, each report is one of four types: (1) a high-credibility, preference-consistent report, (2) a high-credibility, preference-inconsistent report, (3) a low-credibility, preference-consistent report, and (4) a low-credibility, preference-inconsistent report. I expect that participants' viewing time of each of these reports will be influenced by the favorability of the initial information they receive regarding their position. Table 1 presents descriptive statistics of participants' viewing time based on Favorability and the type of analyst report (Credibility, Preference Consistency).

The last column of Table 1 shows that participants in the Favorable condition spent, on average, 66.59 seconds viewing the analyst reports, while those assigned to the Unfavorable condition spent an average of 90.48 seconds. The mixed-design ANOVA model presented in Table 2 shows the effect of Favorability on participants' viewing time to be significant ($F_{1,90} = 6.15$, $p = 0.02$). This result is consistent with the findings in psychology that individuals spend more time and effort processing information after the receipt of unfavorable information (e.g., Kruglanski 1980, 1990; Ditto and Lopez 1992).

¹³ The results reported in the paper are inferentially and statistically similar to the analysis excluding participants who answered the manipulation check questions incorrectly.

¹⁴ I report all significant differences resulting between the two experimental conditions comprising either of the two Favorability conditions.

TABLE 1
Descriptive Statistics for Viewing Time of Analyst Reports^a

Favorability of Initial Information ^b	High-Credibility Reports		Low-Credibility Reports		Collapsed across Credibility		Collapsed Across Preference Consistency		Collapsed across Credibility and Preference Consistency
	Preference- Consistent Report ^c	Preference- Inconsistent Report ^c	Preference- Consistent Report	Preference- Inconsistent Report	Preference- Consistent Reports	Preference- Inconsistent Reports	High- Credibility Reports	Low- Credibility Reports	
Favorable	18.93 (14.57) n = 44	25.14 (22.22) n = 44	12.55 (14.93) n = 44	9.98 (11.61) n = 44	31.48 (25.03) n = 44	35.11 (25.33) n = 44	44.07 (31.77) n = 44	22.52 (20.55) n = 44	66.59 (44.28) n = 44
Unfavorable	43.02 (45.56) n = 48	16.17 (15.15) n = 48	18.56 (17.43) n = 48	12.73 (16.91) n = 48	61.58 (50.56) n = 48	28.90 (22.64) n = 48	59.19 (44.60) n = 48	31.29 (26.80) n = 48	90.48 (47.82) n = 48

The viewing time is in seconds, with standard deviations presented in parentheses.

^a Participants were given a total of three minutes to view four analyst reports, each containing between 416 and 421 words. Two of the four reports are issued by analysts employed by well-known brokerage firms (high-credibility reports) and two by unknown analysts (low-credibility reports). Within each set, one estimates high earnings and the other estimates low earnings. Participants could view as many or as few of the reports as they wished. The primary dependent measure is the time spent viewing each analyst report.

^b Participants were randomly assigned to choose either a long or a short investment position and were provided either a \$3.76 or a \$3.84 earnings benchmark. Subsequent to the participant making an investment choice between two firms and being told their earnings benchmark, all participants were given the same initial information regarding their chosen investment position (i.e., the consensus analyst forecast of \$3.80). This information was Favorable for those in the long (short) investment position condition who were assigned a \$3.76 (\$3.84) earnings benchmark and was Unfavorable for those in the long (short) investment position condition who were assigned a \$3.84 (\$3.76) earnings benchmark.

^c The reports of analysts forecasting low (high) earnings should be viewed as preference-consistent by participants assigned to the short (long) investment position condition and as preference-inconsistent by those assigned to the long (short) investment position condition.

TABLE 2
Mixed-Design ANOVA for the Viewing Time of Analyst Reports

	df	SS	MS	F	Two-Tailed p-value
Between-Participants Effects:					
Favorability	1	3275.03	3275.03	6.15	0.02
Error	90	47937.65	532.64		
Within-Participant Effects:					
Credibility	1	14028.97	14028.97	27.17	< 0.001
Credibility × Favorability	1	231.44	231.44	0.45	0.51
Error (Credibility)	90	46464.85	516.28		
Preference Consistency	1	4843.65	4843.65	8.47	< 0.01
Preference Consistency × Favorability	1	7572.31	7572.34	13.25	< 0.001
Error (Preference Consistency)	90	51443.12	571.59		
Credibility × Preference Consistency	1	860.96	860.96	2.16	0.15
Credibility × Preference Consistency × Favorability	1	5094.38	5094.38	12.76	< 0.001
Error (Credibility × Preference Consistency)	90	35935.18	399.28		

The four experimental conditions are collapsed into two conditions based on the Favorability of the initial information (i.e., consensus analyst forecast of \$3.80). The Favorable condition is comprised of participants assigned to the long (short) investment position with an earnings benchmark of \$3.76 (\$3.84). The Unfavorable condition is comprised of participants assigned to the long (short) investment position with an earnings benchmark of \$3.84 (\$3.76). I analyze the information using a mixed-design ANOVA model with Favorability as a between-participants variable and the Credibility (high credibility versus low credibility) and Preference Consistency (preference-consistent versus preference-inconsistent) of the analyst reports as two within-participants variables.

Holding constant the amount of preference-consistent information offered by high-credibility and low-credibility sources, H1 predicts that all participants will spend a greater amount of time viewing high-credibility reports than low-credibility reports. Consistent with H1, untabulated results show that participants spent an average of 51.96 seconds viewing reports issued by analysts employed by the two high-credibility equity research firms, compared to an average of 27.10 seconds viewing the reports issued by analysts employed by the two low-credibility firms. This difference leads to the significant main effect of Credibility on participants' viewing time of the four reports ($F_{1,90} = 27.17$, $p < 0.001$), as noted in Table 2.

The mixed-design ANOVA in Table 2 also shows a main effect of Preference Consistency on participants' viewing time of the four reports ($F_{1,90} = 8.47$, $p < 0.01$). This main effect, however, is qualified by a two-way interaction of Preference Consistency and Favorability on information viewing times ($F_{1,90} = 13.25$, $p < 0.001$). Compared to participants in the Favorable condition, participants in the Unfavorable condition allocated more of their viewing time to preference-consistent information than to preference-inconsistent information. Participants in the Favorable condition spent approximately the same amount of time viewing the preference-consistent analyst reports (mean = 31.48 seconds) and preference-inconsistent analyst reports (mean = 35.11 seconds) ($t_{43} = 1.01$, $p = 0.32$). However, participants in the Unfavorable condition spent significantly more time viewing the preference-consistent analyst reports (mean = 61.58 seconds) than the preference-inconsistent analyst reports (mean = 28.90 seconds) ($t_{47} = 3.65$, $p = 0.001$).

Credibility and Preference Consistency

Although there is a significant main effect of Credibility on the amount of time participants spent viewing the analyst reports, this main effect is qualified by a significant three-way interac-

tion of Credibility, Preference Consistency, and Favorability ($F_{1,90} = 12.76$, $p < 0.001$). The significant interaction suggests that participants in the Favorable and Unfavorable conditions allocate their viewing time differently across the four analyst reports based on source credibility and the preference-consistency of the report. That is, the two-way interaction of Credibility and Preference Consistency on participants' viewing time varies across the two Favorability conditions. Panels A and B in Figure 2 show graphical presentations of the average viewing times of each of the four analyst reports for participants in the Favorable Condition and Unfavorable Condition, respectively.

Hypothesis 2a predicts that participants in the Favorable condition will be balanced in their acquisition of high-credibility information, choosing to view an equal amount of preference-consistent and preference-inconsistent information. However, H2b predicts that participants in the Unfavorable condition will prefer to view more high-credibility, preference-consistent information than high-credibility, preference-inconsistent information.

Simple effect tests indicate that participants in the Favorable condition spent *more* time viewing the high-credibility, *preference-inconsistent* analyst report (mean = 25.14 seconds) than the high-credibility, preference-consistent analyst report (mean = 18.93 seconds) ($t_{43} = 2.05$, $p = 0.05$). Although this finding does not support H2a, it does not necessarily suggest that participants receiving favorable information regarding their investment position lack objectivity in their acquisition of additional information. Instead, this result is redolent of investors' openness to credible, contrary information (Festinger 1957, 1964; Frey 1986). Additional analysis indicates that participants in the Favorable condition spent equal time viewing the preference-consistent (mean = 12.55 seconds) and preference-inconsistent reports (mean = 9.98 seconds) issued by low-credibility analysts ($t_{43} = 1.00$, $p = 0.33$).

Turning next to participants assigned to the Unfavorable condition, I find that with both high-credibility and low-credibility analyst reports, participants preferred to view the preference-consistent report over the preference-inconsistent report. Specifically, participants in the Unfavorable condition spent, on average, 43.02 seconds viewing the high-credibility, preference-consistent analyst report compared to an average 16.17 seconds viewing the high-credibility, preference-inconsistent analyst report ($t_{47} = 3.63$, $p < 0.001$). This result supports H2b. Similarly, these participants spent, on average, 18.56 seconds viewing the low-credibility, preference-consistent report compared to an average 12.73 seconds viewing the low-credibility, preference-inconsistent report ($t_{47} = 1.88$, $p = 0.07$).

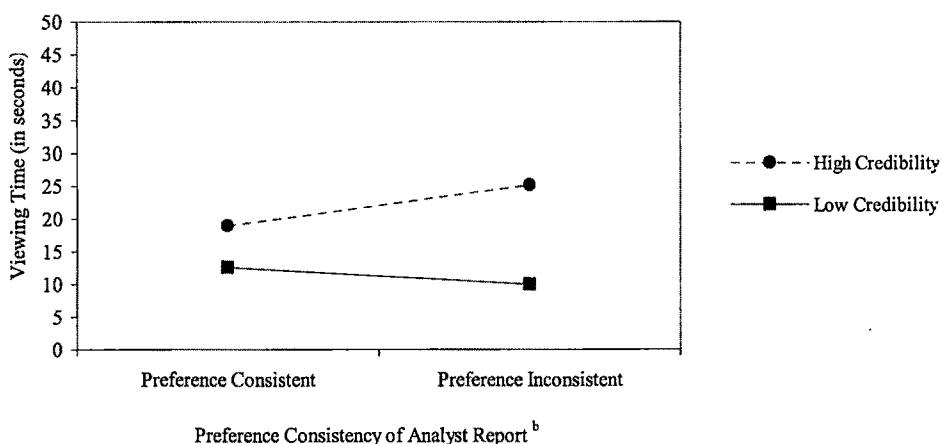
Credibility versus Preference Consistency

Results indicate that participants in the Unfavorable condition seek high-credibility, preference-consistent information over high-credibility, preference-inconsistent information, consistent with H2b. However, I expect these individuals' need to gather information supporting their position to be great enough to counteract the need to gather high-credibility information. Contrary to findings in psychology (e.g., Frey 1981, 1986), H3 predicts that investors receiving initial information that is unfavorable regarding their position will not acquire more high-credibility, preference-inconsistent information than low-credibility, preference-consistent information.

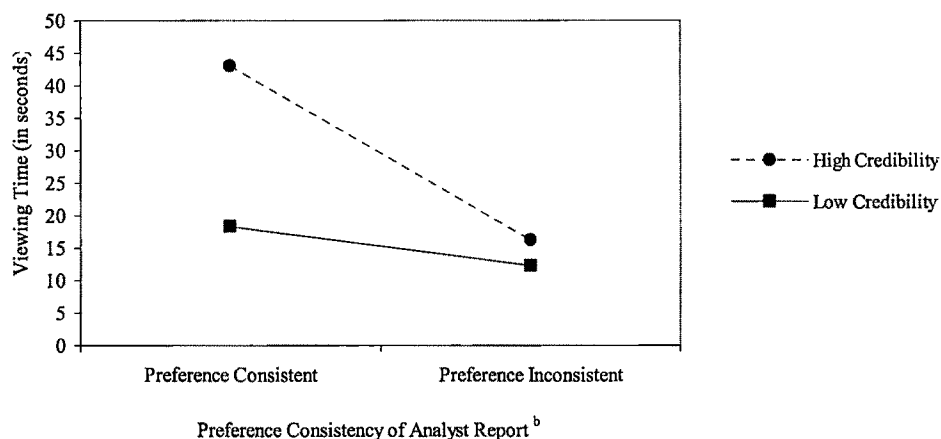
Results support H3. Table 1 shows participants in the Unfavorable condition spent an average of 18.56 seconds viewing the low-credibility, preference-consistent analyst report compared to an average of 16.17 seconds viewing the high-credibility, preference-inconsistent analyst report. The difference in these viewing times is not significantly different ($t_{47} = 0.70$, $p = 0.49$). This result is compelling as previous findings in psychology suggest that individuals receiving unfavorable information prefer additional information that is credible to that which is preference-consistent (e.g., Frey 1981, Ditto et al. 1998). In the current study, however, only participants in the Favorable condition favored credibility over preference-consistency. Putting this result into further

FIGURE 2
Graphical Depiction of Viewing Time of Four Analyst Reports

Panel A: Favorable Condition^a



Panel B: Unfavorable Condition^a



^a The Favorable condition is comprised of participants assigned to the long (short) investment position with an earnings benchmark of \$3.76 (\$3.84). The Unfavorable condition is comprised of participants assigned to the long (short) investment position with an earnings benchmark of \$3.84 (\$3.76).

^b An analyst report is preference-consistent with a participant's assigned long (short) investment position if the reporting analyst forecasts earnings to be above (below) the participant's earnings benchmark. Conversely, an analyst report is preference-inconsistent with a participant's assigned long (short) investment position if the reporting analyst forecasts earnings to be below (above) the participant's earnings benchmark.

context, participants in the Favorable condition spent an average of 12.55 seconds viewing the low-credibility, preference-consistent analyst report compared to an average of 25.14 seconds viewing the high-credibility, preference-inconsistent analyst report. This difference is highly significant ($t_{43} = 3.82$, $p < 0.001$).¹⁵

Forecasted Earnings

In order to examine whether selective information acquisition influences subsequent judgments, participants were assigned the task of forecasting annual earnings-per-share for the subject firm. Participants provided a preliminary EPS forecast after receiving the consensus analyst forecast of \$3.80 but prior to viewing any of the analyst reports. The average preliminary EPS forecast of participants in each of the four conditions was not different from the consensus analyst forecast of \$3.80 (all $p > 0.20$).¹⁶ After having the option to view additional information in the analyst reports, participants provided their final forecasts. Table 3 presents the mean forecasts of participants in each of the four experimental conditions.

The mean forecast of participants assigned to the long investment position was \$3.85, while the mean forecast of participants in the short investment position condition was \$3.78. The difference of \$0.07 is significant ($F_{1,88} = 51.21$, $p < 0.01$). This result replicates the general findings of Hales (2007) that investors' directional preferences regarding an investment position influence their interpretation of information regarding the investment.¹⁷

Although the average EPS forecast of participants in the Unfavorable, Long condition was greater than the average of those in the Favorable, Long condition, the difference is not significant (mean_{Unfavorable, Long} = \$3.86; mean_{Favorable, Long} = \$3.84) ($t_{45} = 1.01$, one-tailed p -value = 0.16). There is a significant difference, however, in the EPS forecasts of participants in the Unfavorable, Short and Favorable, Short conditions ($t_{43} = 1.65$, one-tailed p -value = 0.05), with participants in the Unfavorable, Short condition forecasting EPS to be lower than those in the Favorable, Short condition (mean_{Unfavorable, Short} = \$3.77; mean_{Favorable, Short} = \$3.79).¹⁸

It is interesting to note that while participants in the Unfavorable condition spent more time gathering additional information than did those in the Favorable condition, participants in the

¹⁵ The results are reported in absolute time (seconds). Due to differences in the total amount of time participants in the Favorable and Unfavorable conditions spent viewing the four reports, I do not test differences in absolute time spent viewing individual reports across the two conditions. However, in calculating the percentage of time participants in both conditions allocated to each report, I find that participants in the Favorable condition spent, on average, 16 percent of their viewing time on the low-credibility, preference-consistent report. This is less than the 21 percent of viewing time, on average, allocated to the low-credibility, preference-consistent report by participants in the Unfavorable condition ($t_7 = 1.48$; $p < 0.07$, one-tailed). On the other hand, participants in the Favorable condition spent, on average, 39 percent of their viewing time on the high-credibility, preference-inconsistent report, compared to the 21 percent spent, on average, by participants in the Unfavorable condition. This difference is significant ($t_7 = 3.85$; $p < 0.001$).

¹⁶ This suggests that the pattern of participants' additional information search was not due to differing expectations.

¹⁷ Hales (2007) predicted and found that individuals assigned to hold a long investment position forecasted earnings to be higher than that forecasted by those assigned to hold a short investment position when evaluating information that implied a loss on their respective investments. However, Hales (2007) found that the earnings forecasts of individuals evaluating information that implied a gain on their respective investments were not significantly different. My findings, however, indicate a difference in the estimates of participants in the Favorable, Long condition and those of participants in the Favorable, Short condition ($t_{42} = 3.96$, $p < 0.01$). A possible reason for this difference is the information participants had an opportunity to view. The experimental design included high-credibility sources forecasting annual earnings at both the top and bottom of the range of individual analysts' forecasts. Therefore, high-credibility information was available for those holding a long (short) position to support a higher (lower) earnings forecast.

¹⁸ Forecast errors (i.e., Forecasted EPS - Actual EPS) are presented in Table 3. Actual EPS was \$3.81. Participants in the Unfavorable, Short condition had an average forecast error of -\$0.04, where participants in the Favorable, Short condition had an average forecast error of -\$0.02. The difference in these average forecast errors is significant ($t_{43} = 1.65$, one-tailed p -value = 0.05). Participants in the Unfavorable, Long condition had an average forecast error of \$0.05, where those in the Favorable, Long condition had an average forecast error of \$0.03. The difference in these average forecast errors is not significant ($t_{45} = 1.01$, one-tailed p -value = 0.16).

TABLE 3
Participants' Final Forecasted Earnings-Per-Share

Earnings Benchmark	Investment Position		Collapsed across Investment Position
	Long	Short	
\$3.76	\$3.84 (0.04) ^a \$0.03 ^b n = 23	\$3.77 (0.04) -\$0.04 n = 24	\$3.81 (0.06) n = 47
\$3.84	\$3.86 (0.05) \$0.05 n = 24	\$3.79 (0.04) -\$0.02 n = 21	\$3.83 (0.06) n = 45
Collapsed Across Earnings Benchmark	\$3.85 (0.05) n = 47	\$3.78 (0.04) n = 45	\$3.82 (0.06) n = 92

Shaded cells are Favorable Conditions. Cells not shaded are Unfavorable Conditions.

Participants were given the task of forecasting annual EPS for the subject firm. After receiving background information, including the last three years' earnings information, historical stock prices, and the consensus analyst forecast of annual EPS (\$3.80), participants had an opportunity to view four analyst reports. After viewing this information, participants made their EPS forecast. This task provides an opportunity to examine the differential effects of the information viewed on participants' earnings expectations.

^a The numbers in parentheses represent standard deviations.

^b The numbers in italics represent the average forecast error (i.e., Forecast Error = Forecasted EPS - Actual EPS). Actual EPS was \$3.81.

Unfavorable condition provided final EPS forecasts that were, on average, less accurate than those in the Favorable condition. These findings suggest that, in spite of the additional time participants in the Unfavorable condition spent gathering information, their choice to view additional preference-consistent information over a balanced set of both preference-consistent and preference-inconsistent information led them to overestimate (underestimate) the earnings of the firm in which they held the long (short) position.

Lastly, participants' final EPS forecasts are highly correlated with the percentage of total time spent viewing reports issued by analysts forecasting high earnings relative to the consensus ($\rho = 0.56$, $p < 0.01$) and the percentage of total time spent viewing reports issued by analysts forecasting low earnings relative to the consensus ($\rho = -0.57$, $p < 0.01$). Therefore, participants' acquisition of information, which was influenced by the favorability of initial information regarding their individual investment positions, is consequential to their final EPS forecasts. Regardless of the fact that participants' compensation was largely determined by the accuracy of their forecast, participants' final EPS forecasts were biased in a manner consistent with their preferred outcome.

V. CONCLUSION

This study examines whether an investor's psychological preference for information that supports a recent investment decision dominates his/her economic incentive to gather a balanced set of credible information for purposes of future decision making. I argue that the favorability of initial information an investor receives subsequent to taking an investment position sets the course

of his/her information-acquisition process. Given the psychological need to support a belief in a prior action when that action is called into question (Festinger 1957), I predict that investors receiving unfavorable information regarding a recent investment decision will seek additional information that substantiates their decision, even if it requires them to forfeit a certain amount of credibility in that information.

Results from a web-based experiment support my predictions, as I find that the type of information viewed by participants is a function of the favorability of the initial information they received subsequent to making an investment decision. Participants who initially received favorable information chose to view a more balanced set of preference-consistent and preference-inconsistent information compared to participants receiving unfavorable information. Participants receiving initial unfavorable information spent a majority of their viewing time on preference-consistent information, or information that supported the investment position they chose at the beginning of the case. Finally, the participants receiving initial unfavorable information preferred to view at a similar rate less credible information that was preference-consistent and more credible information that was preference-inconsistent.¹⁹ Consequently, participants' earnings expectations were biased in the direction of the information they chose to view.

While archival studies can explore the impact of investors' decisions through stock returns, these studies cannot easily obtain evidence of the processes that precede individuals' investment decisions. An experimental design allows me to manipulate the type of initial information investors receive subsequent to making an investment decision (i.e., favorable versus unfavorable), while controlling the type of additional information available to them, to examine factors that influence information acquisition. In turn, this study allows more nuanced conclusions about investors' use of information relative to extant archival research.

This study contributes to the accounting and psychology literatures examining individuals' search and use of information in three primary ways. First, the results suggest that an investor who initially receives information that casts doubt on a recent investment decision will not gather a balanced set of credible information, but will seek information that bolsters his/her chosen position. The resulting information set overly represents information supporting his/her position. Additionally, some of the over-represented information may have less credibility than information that is under-represented. The outcome of this information-acquisition process is a set of information that can result in biased estimates and forgone profits.

Second, this study contributes to the accounting literature examining the influence of source credibility on investors' use of information (e.g., Teoh and Wong 1993; Clement and Tse 2003; Gleason and Lee 2003; Mercer 2004; Hutton and Stocken 2009). In general, extant research in this area does not consider the current investment position of the individual seeking information. My study suggests that investors are mindful not only of the credentials of the information source, but also of the source's viewpoint. As such, the results have implications for the generality of findings that investors' reaction to information is often based on source credibility (e.g., Teoh and Wong 1993; Clement and Tse 2003; Gleason and Lee 2003). Results also suggest a potential reason for the finding that the market's pricing of firms with short-selling constraints is less efficient than that for firms with short-sale opportunities (Diamond and Verrecchia 1987; Figlewski and Webb 1993). This inefficiency may stem from bullish investors' lack of attention to bad news.

¹⁹ Although I find a difference in participants' perceived credibility of the available analyst reports, it is possible that their familiarity with the two well-known brokerage firms led them to rate the credibility of those reports as higher than those issued by two unknown firms. However, given the findings that analysts' forecast accuracy is positively associated with employment at large, high-status brokerage firms (e.g., Clement 1999; Hong and Kubik 2003), I do believe that an analyst's association with a particular employer speaks to his/her credibility.

Last, my study contributes to the psychology literature, as it suggests that prior findings of individuals' sensitivity to information credibility, especially when the information is preference-inconsistent, represent a boundary condition (e.g., Lowin 1969; Frey 1981; Ditto et al. 1998). These psychology studies provide experimental participants with information that varies at extreme levels of credibility. The capital markets information environment, on the other hand, offers investors a vast amount of information from sources representing a wide range of credibility. My findings suggest that investors looking to support a previous investment decision are willing to forgo a certain level of source credibility to find supportive information. However, at the point that the variance in two sources' credibility becomes too extreme, investors will likely choose to acquire information from the more credible source.

APPENDIX

INVESTMENT DECISION

After conducting more than 40 hours of research on the pharmaceutical industry, you concluded that two firms in the industry are **undervalued [overvalued]**. Based on your research you would like to take a **long position [short position]** in one of these two stocks.

Note: A **long [short]** investment position is like taking a bet that the stock price will **increase [decrease]**.

Upon comparing the two firms, differences between the firms have been narrowed down to a list of factors. These factors, grouped by firm, are as follows:

Firm X			
Last three years' growth rates in revenue:			
	2006	2005	2004
sales volume	4.80%	5.40%	3.10%
pricing change	1.50%	0.60%	1.00%
currency exchange rate	0.50%	0.70%	3.40%
total growth	6.80%	6.70%	7.50%
Operating Profit Margin last three years:			
	2006	2005	2004
	21%	18%	19%
R&D expense as a % of Total Revenue: 12%			
Resignation of CEO, William Lawson, earlier this year			
Joint ventures with laboratories in UK, France, Germany, India & China. A licensing agreement is also held with one overseas laboratory.			
Increased competition from generic brands in European market			

Firm Y			
Last three years' growth rates in revenue:			
	2006	2005	2004
sales volume	6.30%	7.00%	3.60%
pricing change	1.70%	0.60%	0.60%
currency exchange rate	0.50%	0.50%	4.10%
total growth	7.90%	8.10%	8.30%
Operating Profit Margin last three years:			
	2006	2005	2004
	25%	27%	27%
R&D expense as a % of Total Revenue: 10%			
New independent auditor, Ernst & Young LLP; Previous auditor, PricewaterhouseCoopers			
Manufacturing facilities located in four Western European countries and one South American country.			
2 patent expirations expected in next 18 months; FDA to review application for short extension of patent protection			

All other factors you would consider in making this investment were similar across the two firms, including the prices of the respective stocks.

It is important to remember that part of your payment is based on the performance of the investment position you choose. Please choose the firm in which you would like to take a **long [short]** investment position.

INVESTMENT POSITION: HIGHER [LOWER] EARNINGS MEAN MORE POINTS

You chose a long [short] position in Firm X. The price of Firm X stock at the time of your investment was approximately \$66 per share. Using the firm's forward price-to-earnings (P/E) ratio of 17.55 [17.19], the implied EPS for the upcoming fiscal year end is \$3.76 [\$3.84]. Therefore, your investment benchmark is \$3.76 [\$3.84].

An additional 2 points will be added to your total for every penny by which actual annual EPS is above [below] \$3.76 [\$3.84], up to 25 points.

Your total points will be reduced by 2 points for every penny by which actual annual EPS is below [above] \$3.76 [\$3.84], up to 25 points.

If the firm's actual EPS is \$3.76 [\$3.84], you will neither earn nor lose any points.

In short, higher [lower] earnings leave you with more points. The higher [lower] the firm's relative performance, the better off you are.

REFERENCES

- Bettman, J. R., E. J. Johnson, and J. W. Payne. 1990. A componential analysis of cognitive effort in choice. *Organizational Behavior and Human Decision Processes* 45 (1): 111–139.
- Brehm, J. W. 1956. Post-decision changes in the desirability of alternatives. *The Journal of Abnormal Psychology and Social Psychology* 52: 384–389.
- , and A. R. Cohen. 1962. *Explorations in Cognitive Dissonance*. New York, NY: Wiley.
- Clement, M. B. 1999. Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? *Journal of Accounting and Economics* 27 (3): 285–303.
- , and S. Y. Tse. 2003. Do investors respond to analysts' forecast revisions as if forecast accuracy is all that matters? *The Accounting Review* 78 (1): 227–249.
- Diamond, D. W., and R. E. Verrecchia. 1987. Constraints on short selling and asset price adjustment to private information. *Journal of Financial Economics* 18: 277–311.
- Ditto, P. H., and D. F. Lopez. 1992. Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology* 63: 568–584.
- , G. D. Munro, A. M. Apanovitch, J. A. Scepansky, and L. K. Lockhart. 2003. Spontaneous skepticism: The interplay of motivation and expectation in responses to favorable and unfavorable medical diagnoses. *Personality and Social Psychology Bulletin* 29 (9): 1120–1132.
- , J. A. Scepansky, G. D. Munro, A. M. Apanovitch, and L. K. Lockhart. 1998. Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology* 75 (1): 53–69.
- Erdem, T., and J. Swait. 1998. Brand equity as a signaling phenomenon. *Journal of Consumer Psychology* 7 (2): 131–157.
- Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.
- 1964. *Conflict, Decision, and Dissonance*. Stanford, CA: Stanford University.
- Figlewski, S., and G. P. Webb. 1993. Options, short sales, and market completeness. *The Journal of Finance* 48 (2): 761–777.
- Frey, D. 1981. Postdecisional preferences for decision-relevant information as a function of the competence of its source and the degree of familiarity of this information. *Journal of Experimental Social Psychology* 17: 51–67.
- 1986. Recent research on selective exposure to information. In *Advances in Experimental Social Psychology*, Vol. 19, edited by L. Berkowitz, 41–80. New York, NY: Academic Press.
- , and D. Stahlberg. 1986. Selection of information after receiving more or less reliable self-threatening information. *Personality and Social Psychology Bulletin* 12 (4): 434–441.
- , and R. A. Wicklund. 1978. A clarification of selective exposure. *Journal of Experimental Social Psychology* 14: 132–139.
- Gleason, C. A., and C. M. C. Lee. 2003. Analyst forecast revisions and market price discovery. *The Accounting Review* 78 (1): 193–225.
- Hales, J. 2007. Directional preferences, information processing, and investors' forecasts of earnings. *Journal of Accounting Research* 45 (3): 607–628.
- Hong, H., and J. D. Kubik. 2003. Analyzing the analysts: Career concerns and biased earnings forecasts. *The Journal of Finance* 58 (1): 313–351.
- Hutton, A. P., and P. C. Stocken. 2009. Prior forecasting accuracy and investor reaction to management earnings forecasts. Working paper, Boston College.
- Jennings, R. 1987. Unsystematic security price movements, management earnings forecasts, and revisions in consensus analyst earnings forecasts. *Journal of Accounting Research* 25 (1): 90–110.
- Kruglanski, A. W. 1980. Lay epistemologic—Process and contents: Another look at attribution theory. *Psychological Review* 87 (1): 70–87.
- 1990. Motivations for judging and knowing: Implications for causal attribution. In *The Handbook of*

- Motivation and Cognition: Foundations of Social Behavior*, Vol. 2, edited by E. T. Higgins, and R. M. Sorrentino. 333–368. New York, NY: The Guilford Press.
- , and D. M. Webster. 1996. Motivated closing of the mind: “Seizing” and “freezing.” *Psychological Review* 103 (2): 263–283.
- Leventhal, G. S., and J. W. Brehm. 1962. An experiment on volition of choice. In *Explorations in cognitive dissonance*, edited by Brehm, J. W., and A. R. Cohen. New York, NY: Wiley.
- Lowin, A. 1967. Approach and avoidance: Alternate modes of selective exposure to information. *Journal of Personality and Social Psychology* 6 (1): 1–9.
- 1969. Further evidence for an approach-avoidance interpretation of selective exposure. *Journal of Experimental Social Psychology* 5 (3): 265–271.
- Mercer, M. 2004. How do investors assess the credibility of management disclosures? *Accounting Horizons* 18 (3): 185–196.
- Rogers, J. L., and P. C. Stocken. 2005. Credibility of management forecasts. *The Accounting Review* 80 (4): 1233–1260.
- Sprinkle, G. B. 2000. The effect of incentive contracts on learning and performance. *The Accounting Review* 75 (3): 299–326.
- Stickel, S. E. 1992. Reputation and performance among security analysts. *The Journal of Finance* 47 (5): 1811–1836.
- Teoh, S. H., and T. J. Wong. 1993. Perceived auditor quality and the earnings response coefficient. *The Accounting Review* 68 (2): 346–366.
- Wicklund, R. A., and J. W. Brehm. 1976. *Perspectives on Cognitive Dissonance*. Hillsdale, NJ: Lawrence Erlbaum Associates.

The Spread of Aggressive Corporate Tax Reporting: A Detailed Examination of the Corporate-Owned Life Insurance Shelter

Jennifer L. Brown
Arizona State University

ABSTRACT: This study investigates the spread of aggressive corporate tax reporting by modeling a firm's decision to adopt the corporate-owned life insurance (COLI) shelter. Prior studies identify firm characteristics associated with aggressive tax reporting (Desai and Dharmapala 2006; Frank et al. 2009) and tax shelter participation (Wilson 2009; Lisowsky 2010). This study examines whether social environment factors explain the pattern of tax shelter adoption. Building on theory related to the diffusion of innovations and institutional isomorphism, I hypothesize direct and indirect ties between prior and potential shelter adopters influence the spread of shelter use. I find that network ties via board interlocks increase the likelihood of adopting the COLI shelter. I also find weak evidence that COLI use spreads geographically. However, I find no evidence that the spread of COLI use is concentrated among a particular set of audit firms or industries.

Keywords: *tax shelters; tax aggressiveness; corporate reporting; diffusion; institutional isomorphism; board interlocks.*

Data Availability: *All data are publicly available from sources identified in the text.*

I. INTRODUCTION

During the 1990s, the corporate tax shelter industry boomed, drawing coverage from the financial press and a full-scale crackdown by the U.S. Treasury. In 1998, *Forbes* magazine described a new breed of tax shelters in its cover story, "The Hustling of X-Rated Shelters" (Novack and Saunders 1998). The following year, the U.S. Department of the Treasury (1999) released a 164-page report urging the adoption of numerous legislative measures to curb the growing tax shelter problem. However, to date we have limited evidence on the factors that affect a firm's decision to adopt a tax shelter (Wilson 2009; Lisowsky 2010).

This paper is based on my dissertation, and I appreciate the helpful suggestions and insights of my dissertation committee: John Robinson (chair), Steve Limberg, Linda Krull, Rob Williams, and Jay Hartzell. I also thank Lil Mills, Stephanie Sikes, Steve Kaplan, Katharine Drake, and workshop participants at Arizona State University, Emory University, Florida State University, George Mason University, Rice University, Texas Christian University, University of Illinois at Urbana-Champaign, The University of Texas at Austin, and University of Washington for valuable comments. I gratefully acknowledge the financial support of the Deloitte & Touche Foundation.

Editor's note: Accepted by Thomas Omer.

Submitted: June 2008
Accepted: June 2010
Published Online: January 2011

Like the Enron-era wave of financial accounting scandals, the most recent corporate tax shelter boom alarmed regulators because it signaled a shift in the corporate norm toward aggressive reporting. In outlining its concern over the proliferation of tax shelters, the U.S. Treasury points to the following quote from the New York Bar Association:

The constant promotion of these frequently artificial transactions breeds significant disrespect for the tax system, encouraging responsible corporate taxpayers to expect this type of activity to be the norm, and to follow the lead of other taxpayers who have engaged in tax advantaged transactions.

U.S. Treasury (1999, 3)

This quote motivates the question, “How does a corporate practice such as the use of tax shelters spread?” and suggests that later shelter adopters imitate early shelter adopters because they see tax shelter use as the new norm. I investigate the spread of aggressive corporate tax reporting by examining investments in a specific tax shelter: corporate-owned life insurance (COLI).¹ Drawing on theories from the innovation-diffusion literature, I test whether factors related to a firm’s social environment help to explain tax shelter participation.

Rogers (2003, 11) defines diffusion as “the process by which (1) an *innovation* (2) is *communicated* through certain *channels* (3) *over time* (4) among the members of a *social system*” (emphasis in the original). Because firms’ strategic choices are similar to innovations, researchers have looked to innovation-diffusion theory for insights on how and why certain firm practices spread (such as poison pills [Davis 1991], investor relations departments [Rao and Sivakumar 1999], multi-divisional form [Palmer et al. 1993], golden parachutes [Davis and Greve 1997], options back-dating [Bizjak et al. 2009], and private equity targets [Stuart and Yim 2010]). The decision to adopt an innovation is heavily influenced by the decision-maker’s broader social environment (Rogers 2003). Before adopting an innovation, decision-makers must have knowledge of the new strategy—knowledge often gained through direct contacts between peers or indirect contacts with other agents of change within the social environment. Once aware of the innovation, potential adopters seek to reduce the inherent uncertainty that surrounds the implementation of novel ideas by looking to the experiences of their peers.

The importance of obtaining information on new strategies vicariously by observing prior adopters links diffusion theory with a separate, but related field of study: institutional isomorphism. Institutional isomorphism is concerned with the homogenization of organizational behavior, when, why and how firms mimic one another. When faced with uncertainty, organizations economize on search costs and imitate the behavior of other organizations (Cyert and March 1963). Mimetic, or imitative, organizational change can be thought of as a contagion process that spreads fashionable practices from one firm to another (Haveman 1993). Theory on the diffusion of innovations and institutional isomorphism suggests that firms imitate their peers in the adoption of new strategies (for example, see Haveman 1993; Haunschild 1993; Rao and Sivakumar 1999). I extend the application of innovation-diffusion theory to examine the spread of aggressive tax reporting, focusing on the innovative COLI shelter developed in response to the Tax Reform Act of 1986 (TRA86).

First, I analyze the pattern and timing of COLI shelter adoptions. I find that firms’ adoption decisions appear to be influenced by prior COLI adopters and by the possibility of changes to the tax law. Then, using a matched control sample design, I estimate a cross-sectional logit model of tax shelter adoption that includes four social environment factors suggested by prior studies on the diffusion of strategic corporate practices: (1) direct contacts between prior and potential adopters

¹ I describe the development of the corporate-owned life insurance shelter in Section II. Unless otherwise indicated by context, I use the term “COLI” to refer specifically to the COLI shelter, not corporate-owned life insurance in general.

(proxied by board ties); (2) indirect ties through shared professionals (proxied by auditor ties); and imitation of (3) structurally equivalent (industry) and (4) geographically proximate peers.

I find that direct ties with prior COLI adopters via board interlocks increase the probability that a firm adopts COLI. I also find some evidence that COLI adoptions spread geographically. However, although auditors were targeted in the public press for their role in the tax shelter boom, I find no evidence that COLI adoptions spread via shared auditors. Similarly, I find no evidence that COLI adoptions spread along industry lines. Finally, I test the robustness of these findings using event history analysis and a hazard model that captures both the likelihood of adoption and the rate at which a firm adopts, and find similar results with respect to board interlocks.

This study contributes to recent literature on aggressive corporate tax reporting by examining the spread of a specific tax shelter, COLI. Prior studies have uncovered various firm characteristics associated with aggressive corporate tax reporting (Desai and Dharmapala 2006; Frank et al. 2009) and tax shelter use (Wilson 2009; Lisowsky 2010). Although the results of these studies help to explain cross-sectional variation in aggressive reporting behavior and help researchers identify which firms were likely engaged in aggressive behavior, they do not offer much insight into the wave of aggressive reporting that prompted the U.S. Treasury's report in 1999. Firm behavior is clearly related to firm characteristics, but it is also embedded in larger social structures and culture. This study examines the impact of those social structures and thus contributes to an understanding of the macro-level behavior that took place during the corporate tax shelter boom.

Section II details the development and mechanics of the COLI tax shelter and the identification of COLI adopters examined in this study. Section III discusses COLI in the context of innovation-diffusion theory. Section IV develops a cross-sectional model of COLI adoption and discusses the estimation results. Section V describes the event history analysis and presents the hazard model estimation results. Section VI concludes.

II. BROAD-BASED LEVERAGED COLI

Development of the COLI Shelter

The COLI shelter can be described best as a tax arbitrage transaction: the taxpayer finances the purchase of an asset that produces tax-exempt income, a cash value life insurance policy, using proceeds from a loan that produces tax-deductible interest expense.² The cash value policy includes a death benefit and a savings component, both of which are accorded preferential tax treatment. Unlike other interest income, which is taxed as earned, the interest credited to the cash value of a policy (the inside buildup) is excluded from gross income (IRC §72).³ Moreover, a taxpayer can access the cash value of his/her policy while still preserving the deferral benefit by borrowing against the policy because loans secured by life insurance contracts are not treated as taxable distributions (IRC §72).

Over time, Congress has tried to limit life-insurance-related tax arbitrage by narrowing the statutory definition of life insurance. Concerned about the excessive amount of "key-man" life insurance being purchased by firms, Congress limited interest deductions on COLI borrowings by capping qualified policy loans at \$50,000 per insured life (TRA86). Instead of abandoning the market for tax arbitrage after TRA86, insurance industry entrepreneurs responded by offering a

² Since payouts upon the death of the insured are generally tax-exempt to the beneficiary, the premium expense related to the production of that income is not deductible, hence the need for the arbitrage transaction.

³ Distributions of the cash surrender value are generally treated first as a tax-free recovery of basis, and only result in includible income when the amounts distributed exceed the taxpayer's investment or basis in the policy (IRC §72(e)). Thus, even taxpayers who withdraw the cash surrender value of a policy can still enjoy a significant deferral benefit.

new product: broad-based, leveraged COLI (BBCOLI). BBCOLI uses volume, covering thousands of a company's rank-and-file employees, to make up for the arbitrage opportunity denied by the \$50,000 per insured policy loan cap.⁴

BBCOLI provides a good setting for studying the spread of aggressive corporate tax reporting. Although defining exactly what constitutes aggressive corporate tax reporting is challenging, BBCOLI is generally classified by accounting researchers and legal scholars as a corporate tax shelter.⁵ Furthermore, details from the Camelot case, *In re CM Holdings, Inc.* (254 B.R. 578, D. Del. 2000), reveal exactly when this innovative tax shelter was developed, a critical element in studying the diffusion of a practice. Last, BBCOLI is a good setting because changes to the law in 1996, government court victories, and the implementation of an IRS amnesty/settlement program effectively shut down the COLI shelter, forcing COLI investors to unwind their investments and bringing COLI investors out of hiding.⁶ Thus, COLI represents a chance to identify and study firms that invest in tax shelters, both those that were caught and those that voluntarily came forward.

Identifying COLI Shelter Participants

I use Lexis-Nexis to search 10-Ks, news articles, and press releases for evidence of COLI activity using the following string: (*company or corporate*) w/5 *owned w/5 life w/5 insurance w/10 (tax or taxes)*. This search captures two main groups of firms: (1) those for which involvement in the COLI shelter is revealed *ex post* because of an IRS dispute, IRS amnesty settlement, or other legal dispute, and (2) those that openly disclose their COLI activity in their financial statements during active COLI shelter years.⁷ Owens & Minor is an example of the first group; footnote disclosures in 2000 and 2001 discuss the IRS Notice the firm received regarding COLI and the reserve charge booked in light of the IRS's COLI court victories. Marriott is an example of the second group, as shown in the following excerpts from the firm's 1995 annual report. The top excerpt comes from Management's Discussion and Analysis (MD&A), and the bottom excerpt is the reconciliation of the firm's effective tax rate (ETR) to the statutory tax rate provided in the notes to the financial statements.

The Company's effective income tax rate declined to 40.0 percent from 41.5 percent in the preceding year, despite expiration of federal jobs tax credit programs, due to the impact of the Company's **corporate-owned life insurance** program and certain other investments.

⁴ Winn-Dixie covered nearly all 36,000 of its employees while Wal-Mart allegedly covered 350,000 of its employees (*Winn-Dixie Stores*, 113 T. C. 54 (1999); *Rice v. Wal-Mart Stores, Inc.* 12 F. Supp. 2d 1207).

⁵ COLI is cited as a classic example of the recent breed of corporate tax shelters (Bankman 2004; Eustice 2002; Gergen 2002) and is included in analyses by Graham and Tucker (2006), Wilson (2009), and Lisowsky (2010).

⁶ In 1996, Congress further restricted COLI arbitrage, limiting a corporation's interest deductions to those on policy loans for a maximum of 20 individuals (Health Insurance Portability and Accountability Act, HIPAA 1996). The IRS successfully litigated three COLI cases, *Winn-Dixie* (113 T. C. 254), *C.M. Holdings* (254 B.R. 578), and *AEP* (136 F. Supp. 2d 762), and implemented an amnesty/settlement program for COLI transactions.

⁷ Tax shelters are usually characterized by secrecy and using a sample of firms that openly disclose their activities in their financial statements is likely to call into question whether these firms truly represent typical tax shelter users. However, Bankman (2002) suggests that firms that think their shelter activities would need a cloak are less likely to engage in tax shelters, while those that decide to adopt tax shelters have convinced themselves that their shelter activity meets a littoral reading of the tax law.

	1995	1994	1993
U.S. statutory tax rate	35.0%	35.0%	35.0%
State income taxes, net of U.S. tax benefit	5.0	5.0	6.0
Corporate-owned life insurance	(1.2)	—	—
Tax credits	(1.4)	(0.7)	(0.8)
Other, Net	2.6	2.2	2.0
	40.0%	41.5%	42.2%

My initial text-string search results in 213 possible COLI adopters. I exclude 28 life insurance companies because these firms serve as the other party in the COLI tax shelter transaction. Bank-owned life insurance, BOLI, is similar to COLI, but since the legislative history surrounding limitations on BOLI is different, I exclude 32 banks from the sample of COLI shelter participants.⁸ After excluding banks and insurance firms, 153 potential COLI shelter participants remain.

Firms often purchase COLI for legitimate business purposes, including insuring against the loss of key executives and providing a perk for outside directors. To ensure that I only include firms that participate in the tax shelter variety of COLI, I examine the remaining 153 firms in detail to confirm participation in BBCOLI. I assume that firms whose COLI programs give rise to tax benefits material enough to warrant a line-item on the tax rate reconciliation are involved in BBCOLI and include these firms in the shelter sample. I identify additional COLI shelter participants based on news articles, court cases, and financial statement details regarding COLI-related IRS disputes or settlements. After examining the 153 firms from my initial search in detail, I identify 43 firms as COLI shelter participants.⁹

Estimate of COLI Tax Savings

Table 1 provides various estimates of the magnitude of COLI tax savings for the sample firms. Twenty-seven (62.8 percent) of the sample firms provide detailed information on their COLI tax savings through their tax footnote rate reconciliations. Using these details, I report estimates of the average annual and total tax savings under the columns labeled “Rate Reconciliation.”¹⁰ On average, over the duration of COLI use, firms report saving \$26 million dollars in taxes, equivalent to 2.9 percent of pre-tax book income.

For firms that disclose the dollar amount of their COLI-related IRS settlement, I report the amount and year of settlement under the columns headed, “Settlement.” The mean (median) settlement amount is \$50.8 million (\$38 million), nearly twice the estimated amount of tax savings reported in firms’ rate reconciliations. The last two columns report estimated COLI savings derived from a variety of other sources, including court cases and footnote disclosures of COLI-related reserve charges. The mean COLI tax savings estimated from these sources is \$46.8 million. In 2000, the staff of the Joint Committee on Taxation estimated that there were 100 COLI cases involving nearly \$6 billion dollars in taxes (Paull 2000). Multiplying the highest average tax savings estimate, \$50.8 million, times the 43 firms in the sample yields an aggregate tax savings estimate of more than \$2 billion, suggesting my sample accounts for roughly one-third of the COLI tax shelter market.

⁸ Neither the economic sham arguments used in court to defeat leveraged COLI nor the 1996 HIPAA restrictions on leveraged COLI directly apply to BOLI because banks generally do not borrow against the policies to fund the premium payments.

⁹ My sample includes all firms whose shelter activity is observable. Tax shelter activity is observable if either (1) the firm is a disclosing firm, (2) the firm is a non-disclosing firm whose activity is revealed *ex post* when detected by regulators, or (3) the firm is a non-disclosing firm whose activity is revealed *ex post* when the firm comes forward for the COLI amnesty program. Tax shelter activity is generally not observable for non-disclosing firms whose activities are not detected by regulators.

¹⁰ Because the dollar-effect of rate reconciliation line items is harder to interpret in years when pre-tax book income is negative, I only include COLI tax savings amounts for years with positive pre-tax book income.

TABLE 1
Estimates of COLI Tax Savings

Obs.	Company	Rate Reconciliation				Settlement		Other	
		Year of Adoption	No. of Years	Annual Savings	Total Savings	Total as a Percent of PTBI	Year Disclosed	Amount	Source
1	Advanced Telecommunications	1991	1	\$585	\$585	1.96%			
2	American Electric Power	1990	7	19,418	135,926	2.49%		\$25,000	136 F. Supp 2d 762 ^a
3	American Greetings	1990	10	9,208	92,085	4.50%		143,000	2001 Q4 reserve charge
4	Ball Corp.	1987	11	3,616	39,775	4.38%			
5	Bassett Furniture Industries	1994	5	726	3,631	2.70%		8,000	1998 fn.: total tax effect
6	Brush Engineered Materials	1990	8	715	5,722	3.33%			
7	Carpenter Technology Corp.	1994	NA						
8	Charming Shoppes Inc.	1993					\$18,477	\$2,004	
9	Commercial Intertech	1992	7	1,118	7,824	3.48%			
10	Diebold Inc.	1987	12	2,428	29,131	2.78%	16,040	2002	
11	Donnelly Corp.	1990	9	514	4,628	4.72%	217,000	2002*	1999 fn.: total tax effect
12	Dow Chemical	1989						11,247	250 F. Supp 2d 748
13	Eaton Corp.	1993	NA						
14	Fina Inc.	1989	9	3,043	27,385	2.23%			
15	GATX Corp.	1988	10	4,007	40,071	3.27%			
16	Goodrich Corp.	1990	7	1,565	10,956	0.91%			
17	Grace (W.R.) & Co.	1989					57,500	2005	
18	Great Plains Energy Inc.	1994					12,700	2002*	
19	Hershey Co.	1989	5	3,043	15,216	0.79%		61,200	2000 fn.: payment to IRS
20	Hillenbrand Industries	1990	NA						
21	Hovnanian Entrprs Inc.	1993	6	675	4,050	2.20%			
22	Hubbell Inc.	1993	NA						
23	IKON Office/AICO Standard	1994					38,000	2001*	2001 reserve charge
24	Manor Care Inc.	1993	5	4,254	21,268	5.69%		10,000	
25	Marriott International Inc.	1995	4	4,438	17,753	0.85%			

(continued on next page)

TABLE 1 (continued)

Obs.	Company	Rate Reconciliation				Settlement		Other	
		Year of Adoption	No. of Years	Annual Savings	Total Savings	Total as a Percent of PTBI	Amount	Year Disclosed	Amount
26	Media General	1990	9	1,525	13,727	1.82%			
27	Miller (Herman) Inc.	1994	3	1,867	5,602	2.80%	22,200	2003*	
28	Norfolk Southern Corp.	1989	10	10,300	103,000	1.15%			
29	OLIN Corp.	1992					41,000	2004*	
30	Owens & Minor Inc.	1995							7,200 2001 reserve charge
31	Pitney Bowes Inc.	1994							76,000 2005/2006 reserve charge
32	PPG Industries Inc.	1994	5	7,249	36,245	1.02%			
33	Rohr Inc.	1994	NA						
34	Ruddick Corp.	1994	3	223	669	3.34%			
35	Sherwin-Williams Co.	1993	7	2,648	18,535	3.85%	20,000	2001*	
36	Sonoco Products Co.	1993	NA						23,300 2000/2001 reserve charge
37	Stride Rite Corp.	1993	5	5,604	28,019	2.05%			
38	Swank Inc.	1986	5	1,046	5,228	2.63%			
39	Trans World Entmt. Corp.	1991	5	326	1,632	9.64%			5,534 254 B.R. 578
40	Unifirst Corp.	1994	5	716	3,579	1.82%			5,500 1999 reserve charge
41	Wal-Mart Stores Inc.	1993	NA						
42	Winn-Dixie Stores Inc.	1993	5	6,124	30,619	1.77%	52,000	2003*	1,599 113 T. C. 254 ^b
43	Xcel Energy Inc.	1993					64,400	2007*	
	MEAN			3,592	26,032	2.90%	50,847		46,832
	MEDIAN			2,428	15,216	2.63%	38,000		17,274

(continued on next page)

TABLE 1 (continued)

* Indicates that the dollar amount of year disclosed includes interest on the tax deficiency.

^a Amount is the tax deficiency for the 1996 tax year only.

^b Amount is the tax deficiency for the 1993 tax year only.

This table presents estimates of COLI-generated tax savings based on details from court cases and firms' financial statements. No. of Years equals the number of years prior to 1998 with positive pre-tax book income (PTBI) for which the firm reports COLI savings on the rate reconciliation. Total Savings equals actual annual savings reported on the rate reconciliation summed across those years. Annual Savings equals total savings divided by No. of Years. Total as a Percent of PTBI equals total savings divided by the sum of PTBI across related years. Settlement amounts are identified from financial statement disclosures. The final two columns present estimates of COLI tax savings from a variety of other sources. All dollar amounts are in thousands.

III. THE DIFFUSION OF COLI

Pattern of COLI Adoption

I examine the pattern of COLI adoptions over time. There are three basic models of diffusion: the internal influence model, the external influence model, and the mixed influence model, which allows for both internal and external influences (Mahajan and Peterson 1985).¹¹ Imitation, suggested by the U.S. Treasury (1999) as a factor in the spread of shelter use, is characteristic of the internal influence model, which predicts that the rate at which an innovation is adopted is a function of the cumulative number of prior adopters:

$$\frac{dN(t)}{dt} = x(t) = fn(N(t-1)), \quad (1)$$

where $N(t)$ is the cumulative number of adopters at time t , and $x(t)$ is the number of new adopters at time t . Under the internal-influence model, the rate of adoption, $\frac{dN(t)}{dt}$, plotted against time produces a bell-shaped curve and the cumulative diffusion curve is S-shaped (Mahajan and Peterson 1985; Rogers 2003).

In contrast, external models assume no relationship between $x(t)$ and $N(t-1)$ —adoption is driven solely by an external source like mass media or a change agent. Under the external-influence model, the cumulative number of adopters increases over time, but at a (constant) decreasing rate (Mahajan and Peterson 1985). COLI was developed by an insurance entrepreneur and marketed by insurance brokers, and these tax shelter promoters, acting as external agents of change, obviously had an impact on the spread of COLI use. Although court documents reveal details of the relationship between shelter promoters and a few COLI adopters, the extent of contact between tax shelter promoters and all potential COLI adopters is unobservable. Therefore, I generally assume that the influence of external change agents, in the form of tax shelter promoters, is constant across the sample of potential COLI adopters examined in this study.¹²

I examine the pattern of COLI adoption by using data from firms' rate reconciliations and disclosures regarding IRS disputes and settlements to estimate t , the year of COLI adoption.¹³ As Figure 1 shows, the cumulative number of COLI adopters generally increases from 1986 to 1990, levels off in 1991 and 1992, and increases again in 1993. The increases in new adoptions in 1990 and 1993 can be explained by the legislative history of anti-COLI measures. By 1990, Congress and the U.S. Treasury were aware of COLI abuses (U.S. Department of the Treasury 1990), and in 1991 anti-COLI legislation was proposed in the Senate. Fearing new restrictions on COLI use, but expecting that any existing COLI programs would be grandfathered, firms that were considering adoption in early 1990 rushed to complete their COLI deals (Gergen 2002). After Congress failed to enact the proposed restrictions in 1991 and 1992, the rate of COLI adoptions surged in 1993.

While Figure 1 indicates that the rate of COLI adoptions over time is likely related to the history of proposed legislative measures to curb COLI use, the observed rate of COLI adoptions

¹¹ Importantly, these models assume that an innovation is legal and do not consider the role of an external regulatory body with the power to curb or halt the rate of adoption. Gergen (2002) analytically models how the risk of detection can incentivize tax shelter promoters and tax shelter users to limit the volume of a shelter's distribution. However, this study is the first to empirically model the spread of shelter use.

¹² Shelter promoters could have systematically targeted firms with the same firm characteristics I model as predictors of adoption, giving rise to an endogeneity problem. However, a systematic bias in the type of firm targeted by shelter promoters should bias against findings related to the social environment factors included in the model.

¹³ Since managers can exercise considerable discretion in the amount of detail disclosed in the firm's rate reconciliation, using rate reconciliations may lead to mismeasurement of t , but any mismeasurement should bias against finding results.

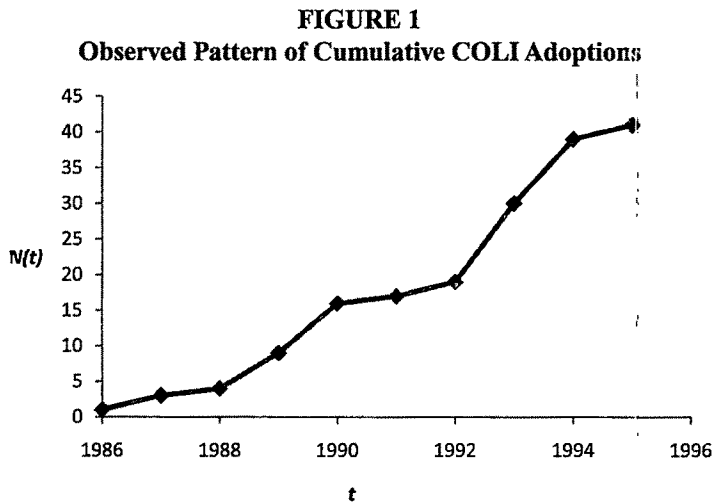


Figure 1 presents the pattern of cumulative COLI adoptions, $N(t)$, observed in the COLI sample.

does not map clearly into the pattern predicted by either a pure internal or pure external source diffusion model. Below, I present four possible mechanisms of diffusion and test whether each helps to explain the spread of COLI use.

Sources of Internal Influence

Innovation-diffusion is linked to mimetic isomorphism, the tendency of firms to imitate one another, because, in an attempt to gain information and reduce the inherent uncertainty surrounding an innovation, potential adopters rely heavily on the experiences of prior adopters. The importance of evaluating an innovation vicariously through the trials of prior adopters is consistent with anecdotes from shelter participants reported by Bankman (1999, 1781): "Most companies are reluctant to be the first purchasers, preferring to purchase a shelter that has been vetted by others." Under frequency-based imitation, as more and more firms adopt a practice, the practice gains legitimacy and eventually becomes taken for granted (March 1981).

Outcome-imitation theorizes that firms adopt strategies when they observe improvements in the performance of peer firms using those strategies (Haunschild and Miner 1997). Abrahamson and Rosenkopf (1993) posit that pressures on firms arising from the threat of lost competitive advantage produce competitive bandwagons, even when the returns to an innovation are unclear. As the number of adopters in a group increases, non-adopters face the risk of falling farther below average if the innovation succeeds. Outcome-imitation is consistent with the U.S. Treasury's (1999) report, which identifies pressure to keep ETRs low and in line with competitors as a driving force behind the tax shelter boom. Both frequency-based imitation and outcome-based imitation imply that the prevalence of a practice within a community will impact the rate of adoption.

Cohesion

Cohesion models of imitation focus on personal contacts between prior adopters and potential adopters. In order to adopt an innovative practice, potential adopters must be aware of the innovation's existence *and* be persuaded to adopt it. A potential adopter can learn about a tax shelter

through its network contacts. More importantly, network contacts with prior adopters can help potential adopters evaluate the benefits and risks associated with a shelter, and therefore impact the persuasion stage of the innovation-diffusion process. Unfortunately, network contacts are often difficult to identify and measure. Tax shelter participants can be connected through common professional association memberships, common board members, common legal advisors, common tax consultants, common auditors, or others.

Board interlocks have been shown to affect the diffusion of a variety of corporate practices (e.g., Davis 1991; Mizruchi 1992; Haunschild 1993; Palmer et al. 1993; Bizjak et al. 2009). The estimates provided in Table 1 suggest that COLI tax savings were significant, and accounts of the negotiations between insurance brokers and firms surrounding the initiation of COLI programs document top management's involvement in the decision to adopt COLI.¹⁴ As such, it is reasonable to assume that board members were aware of COLI and could have influenced the rate of COLI adoptions through their ties to other boards.¹⁵ Even absent this direct type of influence, board interlocks proxy as a general indicator that two firms are socially connected (Strang and Soule 1998). Therefore, I hypothesize:

H1: Ties to prior COLI adopters via board interlocks will increase the likelihood that a firm adopts the COLI shelter.

Professionals, like lawyers, consultants, and accountants, can act as external change agents, influencing the adoption of a new corporate practice by bringing the innovation to their clients. For example, court documents indicate that AEP initially considered COLI at the suggestion of their accountants, Deloitte, Haskins, and Sells (*AEP, Inc. v. U.S.*, 136 F. Supp. 2d 762). Professionals can also play a role in firms' mimetic behavior, acting as the indirect tie connecting prior and potential adopters (DiMaggio and Powell 1983). Haunschild (1994) finds that professional firms act as a conduit for the spread of acquisition premium information. The uncertainty engendered with considering an activity of questionable legitimacy, like tax sheltering, can make firms more likely to rely on information gathered through professional firms, if only to legitimate and rationalize the decision to engage in the activity (Pfeffer 1981).¹⁶

Regulators have focused on the role of public accountants in the spread of the tax shelter industry, criticizing that when a firm advises its audit client on a tax shelter transaction, the firm essentially audits its own work (U.S. Senate Permanent Subcommittee on Investigations 2005). Because regulators investigating the tax shelter industry have placed a spotlight on auditors and because auditors are a natural information conduit between prior and potential shelter adopters, I hypothesize:

H2: Ties to prior COLI adopters via shared auditors will increase the likelihood that a firm adopts the COLI shelter.

Structural Equivalence

In contrast to cohesion models of imitation, structural equivalence models suggest that firms imitate other organizations that have similar relationships with their environment, even in the absence of direct contacts (Galaskiewicz and Burt 1991). Structurally equivalent firms show similar adoption patterns because, in an attempt to maintain their competitive positions, firms

¹⁴ See *AEP*, 136 F.Supp. 2d 762; *In re CM Holdings, Inc.*, 254 B.R. 578; *Winn-Dixie*, 113 T. C. 254 (1999); *Dow Chemical*, 250 F.Supp. 2d 748.

¹⁵ AEP's CFO consulted with the Chairman and CEO of the firm before adopting the COLI program and appeared before the board of directors to explain program once it had been adopted (*AEP, Inc. v. United States*, 136 F.Supp. 2d 762).

¹⁶ When evaluating whether to proceed with a broad-based, leveraged COLI plan, Camelot consulted with Ernst & Young (*In re CM Holdings, Inc.*, 254 B.R. 578).

imitate their successful peers (Burt 1987). Anecdotal evidence suggests that firms face significant pressure to imitate the tax savings strategies of their peers (Novack and Saunders 1998). Table 1 identifies 27 firms that recognized COLI tax savings as a rate reconciliation line-item in the 10-K tax footnotes, suggesting that even though tax shelters are typically characterized by secrecy, knowledge of COLI and its effect on ETR were reasonably public knowledge. Therefore, using industry membership as a proxy for structural equivalence (Fligstein 1985, 1990), I hypothesize:

H3: The presence of prior COLI shelter adopters in a firm's industry will increase the likelihood that a firm adopts the COLI shelter.

Spatial Proximity

Diffusion studies commonly find that spatially proximate actors influence each other (Rogers 2003). For example, Davis and Greve (1997) find that golden parachutes diffused via local business communities, and Burns and Wholey (1993) find regional differences in the adoption of matrix management.¹⁷ Shared membership in a geographic region affords a high level of inter-firm interaction, resulting in the adoption by one firm in an area spurring others to do the same (Strang and Soule 1998). Using regional classifications developed by the Department of Commerce's Bureau of Economic Analysis (BEA region), I hypothesize:

H4a: The presence of prior COLI adopters in a firm's BEA region will increase the likelihood that a firm adopts the COLI shelter.

Regional differences in diffusion rates can also result from the geographic clustering of innovators. Local business elites are connected through a range of formal and informal institutions that facilitate communication, from the country club to local charity organizations. Consequently, those locales closest to an innovation's origin are likely sources of an innovation's earliest adopters (Rogers 2003). Three COLI court cases reveal the involvement of various insurance companies based in New York and New Jersey.¹⁸ Therefore, I assume that the Mid-East BEA region is the center of COLI innovation and hypothesize:

H4b: The greater the geographic distance between the firm's headquarters and the Mid-East BEA region, the less likely the firm is to adopt the COLI shelter.

IV. MODELING COLI ADOPTION

Control Sample

To test whether imitation contributes to the spread of shelter use, I model a firm's decision to adopt COLI using a matched-control design. The foremost factors affecting whether a firm adopts or does not adopt a particular tax shelter are the presence of taxable income to shelter and compatibility between the technology of the shelter and the firm's organizational structure and asset and income mix.¹⁹ The technology of COLI is based on maximizing the tax preferences related to insurance by investing in contracts on a large number of employees. To match COLI adopters to peer firms that could have benefited from adopting COLI but chose not to adopt, I form

¹⁷ A parallel stream of literature in political science examines the regional diffusion of state policy innovations. See Berry and Berry (1990) for a review of this literature.

¹⁸ Mutual Benefit (MBL) of New Jersey underwrote the policies purchased by Camelot Music and AEP, which was also approached with a deal funded by New York Life. AIG of New York underwrote the policies purchased by Winn-Dixie.

¹⁹ Industry is one possible way to control for differences in shelter use arising from differences in the match between firms' operations and the technology of a specific shelter; however, to test whether COLI use spread along industry lines, I do not match COLI and control firms based on industry.

a control sample based on profitability and workforce size because these two firm characteristics are likely prerequisites for COLI adoption. First, I rank all Compustat firms into 20 groups based on *ROA* and *lnEMP* separately, where *ROA* is pre-tax book income minus minority interest divided by total assets $((PI - MII)/AT)$, and *lnEMP* is the log of the number of employees (EMP).²⁰ Then, I identify the *ROA* and *lnEMP* ranks of each COLI firm in year $t-1$ and include any firms with matching *ROA* and *lnEMP* ranks in that year in the control sample. The resulting one-to-many control sample includes 41 COLI firms and 330 control firms.²¹

To make the collection of board interlock data manageable, I develop a one-to-one control sample from the one-to-many sample and use this for my primary tests. For the one-to-one control sample, I choose the control firm with proxy statement data available on Lexis-Nexis that most closely matches each COLI firm's *ROA* and *lnEMP* in $t-1$. Given the difficulty in identifying shelter users, my control sample likely includes firms that used COLI, but whose sheltering activities have not been detected or disclosed.²²

Cross-Sectional Model of Firms' Decisions to Adopt COLI

Initially, I employ the following cross-sectional logistic model to test whether social environment factors affect the likelihood of COLI adoption (H1–H4):

$$\begin{aligned} \text{Log}(P(t)/[1 - P(t)]) = & \beta_0 + \beta_1 \text{FOREIGN}_{i,t-1} + \beta_2 \text{RD}_{i,t-1} + \beta_3 \text{INDRANK}_{i,t} + \beta_4 \text{SGRWTH}_{i,t} \\ & + \beta_5 \text{ETR}_{i,t-1} + \beta_6 \text{BODLINK1}_{i,t} + \beta_7 \text{AUDITLINK1}_{i,t} + \beta_8 \text{INDLINK1}_{i,t} \\ & + \beta_9 \text{BEALINK1}_{i,t} + \beta_{10} \text{DISTANCE}_{i,t} + \varepsilon_{i,t}, \end{aligned} \quad (2)$$

where $P(t)$ is the probability of adopting the COLI shelter.

Independent Variables

I test H1 by including *BODLINK1*, an indicator variable equal to 1 if the firm shared a board member with a COLI adopter at anytime between 1986 and 1995. I hand-collect information on each COLI and control firm's board of director membership from proxy statements. I assign each director a unique identifier and determine all board interlocks among the combined set of COLI and control firms. Both COLI-to-COLI interlocks and COLI-to-control interlocks result in *BODLINK1* equal to 1, while control-to-control interlocks do not.

BODLINK1 captures network ties through board interlocks generally. Here, I assume that the social ties and shared knowledge generated through board interlocks are not limited to actual interlock years, but spillover to surrounding years. In a later model specification, I construct a more restrictive measure of the influence of board interlocks that accounts for both the timing of shared board members and the timing of interlocked firms' COLI adoptions. Hypothesis 1 predicts that COLI adoptions spread via interlocking directorates; therefore, I expect the coefficient on *BODLINK1* to be positive.

I examine H2, the impact of indirect ties between prior and potential adopters via auditors, by including *AUDITLINK1*, an indicator variable equal to 1 if the firm's auditor in year t audited a prior COLI adopter. I obtain each firm's auditor during the year from Compustat's Company

²⁰ I require that firms are incorporated in the U.S. and have total assets greater than \$1 million. These requirements apply to references to "all Compustat firms."

²¹ Two of the original 43 COLI firms do not have sufficient data to construct all the explanatory variables included in the regression models.

²² As a precaution, I exclude 19 firms that appear in my initial 213 COLI sample from the control sample even though I cannot confirm that these firms used broad-based, leveraged COLI. Inclusion of COLI firms in the control samples increases the noise in the data and would bias against finding results.

Auditor data set, data code AU. Consistent with H2, I expect a positive coefficient on *AUDITLINK1*.

To test H3, I include *INDLINK1*, an indicator variable equal to 1 if, at time t , a prior COLI adopter exists within the firm's industry. Industries are classified according to Barth et al. (1998). *INDLINK1* varies by firm and by year but does not vary across industry within a given year. Hypothesis 3 predicts that firms imitate their structurally equivalent peers in order to stay competitive; thus, I expect a positive coefficient on *INDLINK1*.

Similar to *INDLINK1*, I test H4a by including *BEALINK1*, an indicator variable equal to 1 if, at time t , a prior COLI adopter exists within the firm's BEA region. Hypothesis 4a posits that firms imitate their geographic neighbors; therefore, I expect a positive coefficient on *BEALINK1*. Hypothesis 4b predicts that shelter use will spread geographically starting from the source of the innovation. To test H4b, I construct *DISTANCE*, an ordinal variable equal to 1 for firms headquartered in the Mid-East BEA region; 2 for firms headquartered in the neighboring New England, Southeast, and Great Lakes BEA regions; 3 for firms headquartered in the Plains and Southwest BEA regions; and 4 for firms headquartered in the Rocky Mountain and Far West BEA regions. I expect the coefficient on *DISTANCE* to be negative.

Control Variables

The main control variables in Equation (2), *FOREIGN*, *RD*, *ETR*, and *SGRWTH*, are drawn from two recent studies that explicitly model firm characteristics associated with tax shelter participation (Wilson 2009; Lisowsky 2010). Both studies focus on factors that can be used to infer shelter use rather than factors related to the decision to adopt a shelter, and both studies compare non-shelter firms to a group of shelter firms, wherein the participants of several different types of tax shelters are combined together and treated as one class. One issue with grouping participants from different types of shelters is that each shelter has its own technology, and only firms whose traits match the technology of a particular shelter can use it. By modeling a firm's decision to adopt a single type of shelter, I hold differences in the technologies of various shelters constant. Consequently, I do not necessarily expect the same relationship between the control variables and the decision to adopt the COLI shelter that Wilson (2009) and Lisowsky (2010) hypothesize for shelter participation in general.²³

Wilson (2009) finds that large book-tax differences (BTDs) are a marker of shelter use, and Lisowsky (2010) finds that tax shelter use is positively related to lagged ETR. Shelters that produce permanent book-tax differences are the most sought after (U.S. Treasury 1999; Weisbach 2002) as these reduce the firm's ETR and enable managers to simultaneously report low taxable income and high financial income. In terms of its financial statement effect, the COLI shelter is an archetype of recent corporate tax shelters.²⁴ I include the firm's prior year ETR, ETR_{t-1} , as a control variable.²⁵ However, since firms can adopt COLI either to obtain a competitive ETR or to maintain an already low and competitive ETR when an alternative tax-savings strategy dries up, I make no directional prediction for the relationship between ETR_{t-1} and the likelihood of adopting COLI.

Both Wilson (2009) and Lisowsky (2010) include ROA to capture profitability. Firms with greater profitability have more income to shelter. Since matching on ROA precludes including it as a control variable, I include growth in sales, *SGRWTH*, as an alternative measure of management's

²³ I omit several of the variables used by Wilson (2009) and Lisowsky (2010) because these variables capture the effect of tax shelter use rather than factors affecting the decision to adopt a shelter.

²⁴ See the Appendix for details on the financial statement reporting of COLI transactions.

²⁵ *ETR* is calculated as total income taxes (TXT) divided by pre-tax book income (PI). Following Gupta and Newberry (1997), I set *ETR* equal to 1 if TXT/PI is greater than 100 percent, and I set *ETR* equal to 0 when TXT/PI is negative.

incentive to shelter. I expect firms with higher sales growth to be more likely to adopt a shelter; therefore, I expect a positive relationship between *SGRWTH* and the likelihood of adopting COLI.

Since several corporate tax shelters involve foreign operations and intangibles, Wilson (2009) and Lisowsky (2010) predict a positive relationship between measures of these firm characteristics and tax shelter participation. However, compared to other codified and shelter-related, tax-savings strategies involving specific types of assets, liabilities, income, and expenses, COLI can be used by a broad set of firms. Therefore, I include two variables that proxy for firms' non-COLI tax savings opportunities, *RD* and *FOREIGN*, but I expect a negative, rather than a positive relationship between these firm characteristics and COLI adoption. To proxy for R&D activity, I include an indicator variable, *RD*_{*t*-1}, equal to 1 if the ratio of R&D expense to sales in the prior year is positive, and 0 otherwise. I proxy for the extent of a firm's foreign operations with an indicator variable, *FOREIGN*_{*t*-1}, equal to 1 if the ratio of the absolute value of a firm's foreign pretax income (PIFO) to the absolute value of its total pretax book income (PI – MII) is greater than 10 percent.

Finally, I include one additional variable suggested by diffusion theory, *INDRANK*. Larger organizations are generally more innovative (Rogers 2003). Furthermore, studies find that firms tend to imitate their successful and prestigious peers (Burns and Wholey 1993; Haveman 1993; Han 1994). To capture prestige, I rank all Compustat firms into 20 groups by year, industry, and total assets. I expect larger firms to adopt earlier and smaller firms to wait and adopt later, if at all; thus, I expect a positive coefficient on *INDRANK*.

Pattern of Board Interlocks

Figure 2 depicts all board interlocks within the combined COLI and one-to-one control sample during the time period 1986–1995. Of the 82 sample firms, 43 (52.4 percent) share a board member with another firm in the sample. COLI firms are shown as circles, including the year of COLI adoption, and control firms are shown as squares. The proportion of COLI firms with board ties to another sample firm (either COLI or control) does not significantly differ from the proportion of control firms with board ties to another sample firm (58.5 percent versus 46.3 percent; $z = 1.11$; $p = 0.27$). However, relative to control firms, COLI firms have significantly more interlocks to other COLI firms (48.78 percent versus 17.07 percent; $z = 3.05$; $p < 0.01$). The pattern of board interlocks depicted is consistent with H1, and suggests that COLI adopters are tied to one another via director contacts.

Pattern of COLI Adoption by Auditor, Industry, and Region

Table 2 shows the distribution of COLI adoptions across years, auditors, industries, and BEA regions. Notably, the first four COLI adopters in the sample each had different external auditors and, by 1990, COLI use had spread to clients of all of the former Big 6 auditors as well as BDO Seidman. The first four adopters are also in three different industries, and while over half of new COLI adoptions occur in the last three years, those adoptions only result in the spread of COLI use to two new industries. The pattern of COLI adoptions does not reveal an obvious concentration of early COLI adopters in any particular industry, auditor, or geographic region.

Table 3 provides further detail on the pattern of COLI use by comparing the distribution of COLI versus control firms across auditors, industries, and BEA geographic regions. Panel A shows that the breakdown of firms by auditor is very similar for the COLI and one-to-one control samples. Furthermore, the Big 6 auditors audited more than 95 percent of each of the three samples: COLI, one-to-one control, and one-to-many control. Panel B indicates that the proportion of firms in the textiles and chemicals industries is higher among COLI firms than for either of the

FIGURE 2
Diagram of Board Interlocks

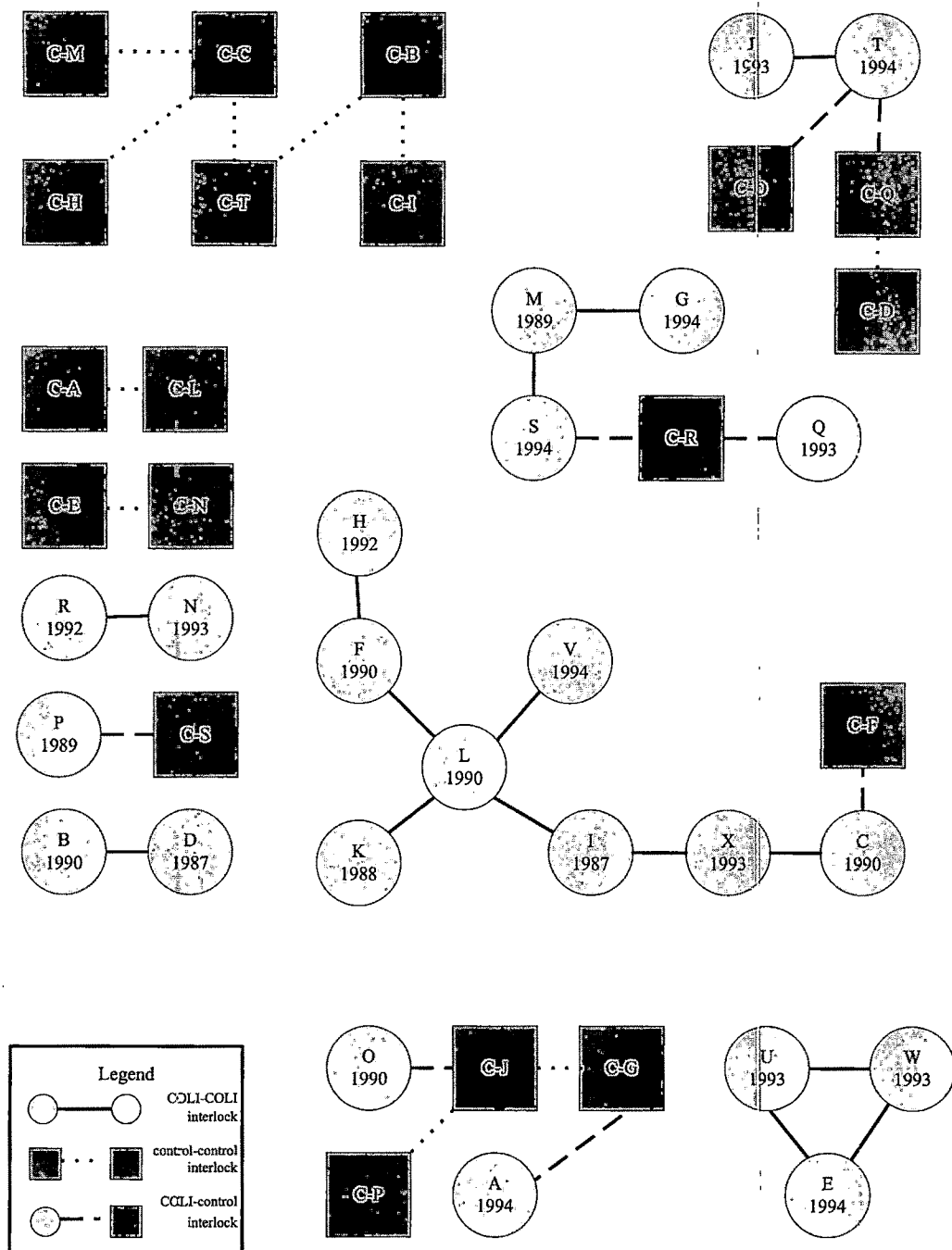


Figure 2 depicts all board interlocks within the combined COLI and control sample ($n = 82$ firms) during the time period 1986–1995. The one-to-one control sample is constructed by matching COLI firms and control firms based on *ROA* and *lnEMP* in the year prior to adoption ($t-1$) and proxy statement data availability.

TABLE 2
Pattern of New COLI Adoptions across Time, Auditors, Industries, and BEA Regions

Year	(1)	(2)	(3)	(4)
	New COLI adopters	New Auditors represented	New Industries represented	New BEA Regions represented
1986	1	1	1	1
1987	2	2	1	2
1988	1	1	1	0
1989	5	2	3	2
1990	7	1	2	0
1991	1	0	1	0
1992	2	0	0	1
1993	11	1	2	1
1994	9	0	0	0
1995	2	0	0	0
Total	41	8	11	7

Sample includes 41 COLI adopters with data to construct all explanatory variables. Column (1) shows the number of new COLI adopters by year. Broad-based, leveraged COLI was developed in 1985 and 1986 in response to restrictions enacted in TRA86 and was effectively shut down by further tax law changes in 1996 (HIPAA). Columns (2)–(4) show the number of new auditors, industries, and BEA regions represented by those adopters. Auditor is determined from Compustat data code AU. Industry classifications are based on Barth et al. (1998). BEA region classifications are obtained from the U.S. Department of Commerce's Bureau of Economic Analysis and assigned based on Compustat data code STATE.

two control samples, but the distribution of firms across industries is otherwise similar among the three groups. Contrary to H2 and H3, the descriptive data in Table 2 and Table 3 show no particular concentration of COLI adopters by auditor or industry.

Panel C of Table 3 shows some difference between the COLI and control samples in the distribution of firms across geographic region, consistent with H4a and H4b. Notably, although 10–12 percent of the control firms are headquartered in the Far West region, there are no COLI firms headquartered there. Also, relative to the one-to-one control sample, there are significantly more COLI firms headquartered in the Great Lakes region ($\chi^2 = 3.36$; $p = 0.07$).

Descriptive Statistics

Table 4 provides descriptive statistics for both the COLI and one-to-one control samples, measured in year t . There is no statistical difference between the COLI and control samples in mean lagged profitability (ROA_{t-1}) or mean lagged workforce size ($\ln EMP_{t-1}$), suggesting an effective matching procedure. Furthermore, there is no statistical difference in the average size of the COLI and control firms. Inconsistent with my expectations, control firms have significantly higher sales growth ($SGRWTH$) relative to COLI firms (12.5 percent versus 7.6 percent; two-tailed $p = 0.04$). Given that the COLI and control firms have similar profitability, one explanation for the lower sales growth among COLI firms is that COLI firms, facing lower growth prospects, focus on maintaining profitability by reducing expenses, specifically reducing tax expense through the use of a tax shelter.

Consistent with H1, tests for differences in proportions indicate that COLI-to-COLI board interlocks are significantly more common than COLI-to-control board interlocks. Contrary to H2, but consistent with the distribution of observations across auditors reported in Table 3, the proportion of COLI firms with links to a prior COLI adopter via shared auditor is actually lower than

TABLE 3
Distribution of COLI and Control Firms across Auditor, Industries, and BEA Regions^a

Panel A: Distribution of Observations across Audit Firms (year *t*)

Auditor ^b	COLI		One-to-One Control Sample		One-to-Many Control Sample	
	Obs.	Percent of Sample	Obs.	Percent of Sample	Obs.	Percent of Sample
Arthur Andersen	5	12.20%	7	17.07%	66	20.00%
Arthur Young	0	0.00%	0	0.00%	2	0.61%
Coopers & Lybrand	5	12.20%	2	4.88%	39	11.82%
Ernst & Young	13	31.69%	12	29.27%	59	17.88%
Deloitte & Touche	4	9.76%	4	9.76%	55	16.67%
KPMG	7	17.07%	8	19.51%	43	13.03%
PricewaterhouseCoopers	5	12.20%	6	14.63%	50	15.15%
Touche Ross	0	0.00%	1	2.44%	4	1.21%
Other	0	0.00%	0	0.00%	8	2.42%
BDO Seidman	1	2.44%	0	0.00%	0	0.00%
Grant Thornton	0	0.00%	0	0.00%	2	0.61%
Kenneth Leventhal	1	2.44%	0	0.00%	1	0.30%
Laventhol and Horwath	0	0.00%	1	2.44%	1	0.30%
Total	41	100.00%	41	100.00%	330	100.00%

(continued on next page)

Panel B: Distribution of Observations across Industries

Industry Description ^c	Primary SIC Codes	COLI		One-to-One Control Sample		One-to-Many Control Sample	
		Obs.	Percent of Sample	Obs.	Percent of Sample	Obs.	Percent of Sample
Mining and construction	1000-1299, 1400-1999	1	2.44%	2	4.88%	13	3.94%
Food	2000-2111	1	2.44%	2	4.88%	21	6.36%
Textiles, printing and publishing	2200-2780	7	17.07%	2	4.88%	29	8.79%
Chemicals	2800-2824, 2840-2899	5	12.20%	2	4.88%	11	3.33%
Pharmaceuticals	2830-2836	0	0.00%	0	0.00%	8	2.42%
Extractive industries	2900-2999,	1	2.44%	1	2.44%	7	2.12%
Durable manufacturers	1300-1399 3000-3569, 3580-3669, 3680-3999	10	24.38%	12	29.25%	78	23.64%
Transportation	4000-4899	2	4.88%	2	4.88%	16	4.85%
Utilities	4900-4999	2	4.88%	0	0.00%	16	4.85%
Retail	5000-5999	3	7.32%	4	9.76%	29	8.79%
Computers	7370-7379, 3570-3579, 3670-3679	7	17.07%	10	24.39%	72	21.82%
Services	7000-7369, 7380-8999	2	4.88%	4	9.76%	30	9.09%
Other	> 9000	0	0.00%	0	0.00%	0	0.00%
Total		41	100.00%	41	100.00%	330	100.00%

(continued on next page)

Panel C: Distribution of Observations across BEA Regions

BEA Region ^d	COLI		One-to-One Control Sample		One-to-Many Control Sample	
	Obs.	Percent of Sample	Obs.	Percent of Sample	Obs.	Percent of Sample
New England Region	4	9.76%	6	14.63%	32	9.70%
Mideast Region	10	24.39%	11	26.83%	65	19.70%
Great Lakes Region	13	31.70%	6	14.63%	70	21.19%
Plains Region	3	7.32%	1	2.44%	25	7.58%
Southeast Region	9	21.95%	7	17.07%	65	19.70%
Southwest Region	1	2.44%	3	7.32%	29	8.79%
Rocky Mountain Region	1	2.44%	2	4.88%	9	2.73%
Far West Region	0	0.00%	5	12.20%	35	10.61%
Total	41	100.00%	41	100.00%	330	100.00%

^a The one-to-many control sample is constructed by ranking all Compustat firms into 20 groups based on *ROA* and *lnEMP* separately, identifying the *ROA* and *lnEMP* ranks of each COLI firm in year $t-1$ and matching control firms based on those *ROA* and *lnEMP* ranks. *ROA* is pre-tax book income minus minority interest divided by total assets ((PI - MII)/AT). *lnEMP* is the log of the number of employees (EMP). The one-to-one control sample is constructed by choosing the closest *ROA* and *lnEMP* match with available proxy statements from the one-to-many control sample.

^b Auditor is determined from Compustat code AU. Firm mergers and acquisitions are reflected above.

^c Industry classifications are based on Barth et al. (1998).

^d BEA region classifications are obtained from the U.S. Department of Commerce's Bureau of Economic Analysis and assigned based on Compustat data code STATE.

TABLE 4
Descriptive Statistics of Regression Variables
(COLI and control firms, year *t*)

Panel A: Distribution of Continuous and Ordinal Variables					Control Sample (n = 41)			
COLI Sample (n = 41)					Mean ^a	Q1	Median	Q3
Variable	Mean	Q1	Median	Q3				
Total Assets (\$billions)	3.020	0.534	1.141	3.096	1.900	0.324	1.075	2.455
<i>LnEMP</i> _{<i>t</i>-1}	2.543	1.581	2.303	3.384	2.408	1.478	2.485	3.045
<i>ROA</i> _{<i>t</i>-1}	0.103	0.064	0.101	0.138	0.100	0.065	0.101	0.130
<i>ETR</i> _{<i>t</i>-1}								
<i>SGR_W</i> <i>TH</i>	0.076	0.018	0.061	0.113	0.125**	0.052	0.090	0.170
<i>INDRANK</i>	8.976	9.000	9.000	10.000	8.780	8.000	9.000	10.000
<i>DISTANCE</i>	1.902	2.000	2.000	2.000	2.171	1.000	2.000	3.000

Panel B: Frequency of Indicator Variables					Control Sample (n = 41)		
Variable	COLI Sample (n = 41)	Control Sample (n = 41)	Difference in Proportions	z-statistic ^b			
<i>FOREIGN</i> _{<i>t</i>-1}	26.83%	41.46%	-14.63%	-1.40			
<i>RD</i> _{<i>t</i>-1}	48.78%	39.02%	9.76%	0.89			
<i>BODLINK</i>	48.78%	17.07%	31.71%	3.05***			
<i>AUDITLINK</i>	80.49%	92.68%	-12.20%	-1.62			
<i>INDLINK</i>	65.85%	75.61%	-9.76%	-0.92			
<i>BEALINK</i>	80.49%	70.73%	9.76%	1.03			

^a *, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels for two-tailed, t-tests of difference in the means.

^b *, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels for two-tailed, tests of differences in proportions. The percentages reported represent the proportion of observations in each subsample where the indicator variable equals 1.

Variable Definitions

*LnEMP*_{*t*-1} = lagged value of the log of the number of employees (EMP);

*ROA*_{*t*-1} = lagged value of pre-tax book income (PTI) minus minority interest (MI), divided by the average total assets (AT);

*ETR*_{*t*-1} = lagged value of total income taxes (TXT) divided by pre-tax book income (PI);

(continued on next page)

$SGRWT$ = sales (SALE) in year t minus sales in year $t-1$, divided by sales in year $t-1$;

$INDRANK$ = decile-ranking resulting from ranking all Compustat firms with assets greater than \$1 million by year, industry, and total assets (AT);

$DISTANCE$ = ordinal variable equal to 1 for firms headquartered in the Mid-East BEA region, 2 for firms headquartered in the neighboring New England, Southeast, and Great Lakes BEA regions, 3 for firms headquartered in the Plains and Southwest BEA regions, and 4 for firms headquartered in the Rocky Mountain and Far West BEA regions;

$FOREIGN_{t-1}$ = indicator variable equal to 1 if the absolute value of foreign pre-tax income (PITO) divided by the absolute value of total pre-tax income (PTI) is greater than 10 percent;

$R\&D_{t-1}$ = indicator variable equal to 1 if R&D expense (XRD) is positive;

$BODLINK$ = indicator variable equal to 1 if the firm has a board interlock with a COLI firm at any time during the period from 1986–1995;

$AUDITLINK$ = indicator variable equal to 1 if the firm's auditor in year t has audited a prior COLI adopter;

$INDLINK$ = indicator variable equal to 1 if, in year t , there is a prior COLI adopter in the firm's industry; and

$BEALINK$ = indicator variable equal to 1 if, in year t , there is a prior COLI adopter in the firm's BEA region.

the proportion of control firms with such links (80.49 percent versus 92.68 percent; $z = -1.62$; two-tailed $p = 0.11$). There is no significant difference between the COLI and control sample in the proportion of firms with links to prior COLI adopters via industry or BEA region.

Cross-Sectional Logit Results

Table 5 reports results for estimating a series of nested cross-sectional logistic regressions constructed from Equation (2). All models have reasonable explanatory power with pseudo- R^2 values ranging from 10.45 percent to 16.33 percent.²⁶ Of the firm-specific characteristics, only $FOREIGN_{t-1}$ is significant in all models. As expected, COLI adopters are less likely to have significant foreign operations relative to non-adopters. Wilson (2009) and Lisowsky (2010) generally find that foreign income is positively associated with shelter use; however, their shelter samples group a variety of shelters together, including some that can only be used by firms with foreign assets or income. As such, the positive association they find between foreign operations and tax shelter use may reflect the technology match between certain shelters and certain firms rather than a determinant of tax shelter adoption generally.

The results for RD_{t-1} and $INDRANK$ are not in the predicted direction and vary across the models. Wilson (2009) also reports mixed results for the effect of R&D intensity on tax shelter use. $INDRANK$ is based on the firm's size (AT) relative to the other firms in its industry. Table 3 and Table 4 show that COLI and control firms are distributed evenly across industries and are not statistically different in size. Furthermore, I match firms based on ROA and $\ln EMP$, diminishing differences in $INDRANK$ across the two groups. The coefficient on ETR_{t-1} is not significant in any of the models. In untabulated results, I substitute ETR_{t-1} with measures of the firm's lagged ETR adjusted for the lagged industry mean (median) ETR, and those measures are also not significant. Since I match firms based on profitability, which is a major determinant of ETR, there is likely little systematic variation in the ETR between the control and COLI firms.²⁷

Consistent with H1, which predicts that board interlocks impact firms' COLI adoption decisions, the coefficient on $BODLINK1$ is positive and significant (1.02; $z = 1.72$; one-tailed $p = 0.04$). Holding all indicator variables equal to 1 and all other variables at their median values, a change in $BODLINK1$ from 0 to 1 is associated with a 73.6 percent increase (from 33.7 percent to 58.5 percent) in the probability that a firm adopts the COLI shelter. $BODLINK1$ is a broad measure that generally proxies for the set of social ties between firms. In the event history analysis presented below, I explore the association between board interlocks and the decision to adopt the COLI shelter further.

Hypothesis 2 predicts that auditors impact firms' COLI adoption decisions. However, the coefficient on $AUDITLINK1$ is negative rather than the predicted positive association; suggesting that prior and potential COLI adopters are not linked via shared auditors. Although the results in Table 5 cannot be generalized to other tax shelters, individual or corporate, COLI use does not appear to have spread through any particular public accounting firm.

Hypothesis 3 predicts the likelihood of adopting COLI increases when there is a prior adopter in a firm's industry. A positive coefficient on $INDLINK1$ would imply that firms imitate their structurally equivalent peers' tax sheltering activities. However, as Table 5 shows, the coefficient on $INDLINK1$ is not statistically significant. One possible explanation is that industry does not accurately capture a firm's structurally equivalent peer group. Another is that, among potential diffusing mechanisms, the spread of COLI is better characterized by a direct-contact cohesion

²⁶ For the least restrictive model, including $BODLINK1$ and $DISTANCE$, the Hosmer and Lemeshow goodness-of-fit χ^2 equals 76.4, with a p-value for the null hypothesis equal to 0.40, indicating that the model is reasonably well fitted.

²⁷ Both COLI and control firms may be engaged in other, non-COLI aggressive tax savings strategies either before, during, or after the COLI adoption decision; these activities could also impact ETR.

TABLE 5
Cross-Sectional Logit Model of COLI Adoption

$$\begin{aligned} \text{Log}(P(t)/[1 - P(t)]) = & \beta_0 + \beta_1 \text{FOREIGN}_{i,t-1} + \beta_2 \text{RD}_{i,t-1} + \beta_3 \text{INDRANK}_{i,t} + \beta_4 \text{SGRWTH}_{i,t} + \beta_5 \text{ETR}_{i,t-1} + \beta_6 \text{BODLINKI}_{i,t} \\ & + \beta_7 \text{AUDITLINKI}_{i,t} + \beta_8 \text{INDLINKI}_{i,t} + \beta_9 \text{BEALINKI}_{i,t} + \beta_{10} \text{DISTANCE}_{i,t} + \varepsilon_{i,t} \end{aligned}$$

Variable	Pred. Sign	Firm Characteristics		H1: BOD Interlock		H2: Auditor		H3: Industry	
		Coeff	z-stat	Coeff	z-stat	Coeff	z-stat	Coeff	z-stat
Intercept	-	-1.958	-1.08	-1.573	-0.85	-1.466	-0.77	-1.980	-1.08
FOREIGN _{t-1}	-	-1.489	-2.36***	-1.205	-1.84**	-1.603	-2.43***	-1.466	-2.32***
RD _{t-1}	-	0.848	1.34	0.585	0.89	0.779	1.20	0.900	1.41
INDRANK	+	0.252	1.20	0.160	0.74	0.398	1.71**	0.303	1.39*
SGRWTH	+	-5.655	-2.08	-4.000	-1.43	-6.672	-2.32	-5.600	-2.08
ETR _{t-1}	?	1.133	0.49	1.044	0.44	0.750	0.31	1.163	0.50
BODLINKI	+			1.020	1.72**				
AUDITLINKI	+					-1.673	-1.95	-0.670	-1.19
INDLINKI	+								
BEALINKI	+								
DISTANCE	-								
Goodness of Fit Statistics									
Pseudo-R ²		10.45%		13.14%		14.24%		11.73%	
Chi-square test		11.88**		14.94**		16.19**		13.33**	
Area under ROC curve		0.70		0.73		0.72		0.71	
H4a: BEA Region									
Variable	Pred. Sign	H4a: BEA Region		H4b: Distance		H1 and H4b			
		Coeff	z-stat	Coeff	z-stat	Coeff	z-stat	Coeff	z-stat
Intercept	-	-2.724	-1.43	-0.693	-0.35	-0.170	-0.08	-0.170	-0.08
FOREIGN _{t-1}	-	-1.615	-2.48***	-1.631	-2.51***	-1.344	-2.00***	-1.344	-2.00***

(continued on next page)

Variable	Pred. Sign	H4a: BEA Region		H4b: Distance		H1 and H4b	
		Coeff	z-stat	Coeff	z-stat	Coeff	z-stat
<i>RD_{t-1}</i>	-	1.113	1.66*	0.883	1.37	0.621	0.92
<i>INDRANK</i>	+	0.242	1.14	0.259	1.19	0.159	0.70
<i>SGR_{WTH}</i>	+	-6.072	-2.08	-6.531	-2.20	-4.687	-1.53
<i>ETR_{t-1}</i>	?	1.441	0.62	0.733	0.31	0.593	0.25
<i>BODLINKI</i>	+					1.110	1.83**
<i>AUDITLINKI</i>	+						
<i>INDLINKI</i>	+						
<i>BEALINKI</i>	+	0.931	1.54*				
<i>DISTANCE</i>	-			-0.529	-1.74**	-0.577	-1.82**
Goodness of Fit Statistics							
Pseudo-R ²		12.62%		13.30%		16.33%	
Chi-square test		14.35**		15.11**		18.56***	
Area under ROC curve		0.70		0.73		0.74	

*, **, *** Indicate significance at the 0.10, 0.05, and 0.01 levels in two-tailed tests (one-tailed when predicted).
The sample includes 41 COLI firms and 41 control firms from a sample matched on ROA and lnEMP in year $t-1$ (the year before adoption).

Variable Definitions:

FOREIGN_{t-1} = indicator variable equal to 1 if the absolute value of foreign pre-tax income (PIFO) divided by the absolute value of total pre-tax income (PI) is greater than 10 percent;

RD_{t-1} = indicator variable equal to 1 if R&D expense (XRD) is positive;

INDRANK = decile-ranking resulting from ranking all Compustat firms by year, industry, and total assets (AT);

SGR_{WTH} = sales (SALE) in year t minus sales in year $t-1$, divided by sales in year $t-1$;

ETR_{t-1} = lagged value of total income taxes (TXT) divided by pre-tax book income (PI);

BODLINKI = indicator variable equal to 1 if the firm has a board interlock with a COLI firm at any time during the period from 1986–1995;

AUDITLINKI = indicator variable equal to 1 if the firm's auditor in year t has audited a prior COLI adopter;

INDLINKI = indicator variable equal to 1 if, in year t , there is a prior COLI adopter in the firm's industry;

BEALINKI = indicator variable equal to 1 if, in year t , there is a prior COLI adopter in the firm's BEA region; and

DISTANCE = ordinal variable equal to 1 for firms headquartered in the Mid-East BEA region, 2 for firms headquartered in the neighboring New England, Southeast, and Great Lakes BEA regions, 3 for firms headquartered in the Plains and Southwest BEA regions, and 4 for firms headquartered in the Rocky Mountain and Far West BEA regions.

model than by an indirect structural equivalence model. Similar to H3, H4a predicts the likelihood of adopting COLI increases when there is a prior adopter in a firm's geographic region. The coefficient on *BEALINK1* is positive, but not significant.

Hypothesis 4b predicts that exposure to COLI through local social networks will be strongest in locales near the large insurance firms underwriting COLI policies; thus, firms with headquarters farther away from the Mid-East BEA region will be less likely to adopt the COLI shelter. Consistent with H4b, the coefficient on *DISTANCE* is negative and significant (-0.529 ; $z = -1.74$; one-tailed $p = 0.04$). The last column in Panel B of Table 5 shows that *BODLINK1* and *DISTANCE* remain statistically significant when they are included in the model together, indicating that COLI adoptions are influenced by two complementary diffusing mechanisms. Holding all indicator variables equal to 1 and all other variables, including *DISTANCE*, at their median values, a change in *BODLINK1* from 0 to 1 is associated with a 84 percent increase (from 31.9 percent to 58.7 percent) in the probability that a firm adopts the COLI shelter. Holding all indicator variables, including *BODLINK1*, equal to 1 and all other variables at their median values, a one-unit change in *DISTANCE*, from *DISTANCE* = 2 to *DISTANCE* = 3, is associated with a 24.5 percent decrease (from 58.7 percent to 44.3 percent) in the probability that a firm adopts the COLI shelter.

The evidence in Table 5 suggests that COLI adoption spread among firms with shared board members and spread out geographically from the Mid-East region. Overall, results from using a one-to-one matched control sample and estimating cross-sectional models based on Equation (2) are consistent with H1 and H4b, but are not consistent with H2, H3, and H4a. In untabulated tests, I use a one-to-many control sample to test H2, H3, H4a, and H4b. The coefficient on *BEALINK1* is positive and significant (0.79 ; $z = 1.86$; one-tailed $p = 0.03$), and the coefficient on *DISTANCE* is negative and significant (-0.50 ; $z = -2.33$; one-tailed $p = 0.01$). The results using a one-to-many control sample are consistent with both H4a and H4b, but are not consistent with H2 and H3.

V. EVENT HISTORY ANALYSIS

Model Specification

The cross-sectional model presented above tests whether certain social environment factors affect the likelihood of adopting COLI at time t , but does not test how those factors affect the timing of COLI adoptions. Below, I follow most diffusion studies and employ an event history or duration model. The variable of interest in a duration model is the length of time, T , that elapses between the time a firm is first "at risk" of adopting a new practice and the time, t , at which the firm adopts the practice. Duration data can be used to estimate a hazard rate, $h(t)$, the rate at which the duration ends in the interval $[t, t + \Delta]$, given that the duration has not ended prior to the beginning of this interval:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t + \Delta t > T \geq t | T \geq t)}{\Delta t} \quad (3)$$

I am interested in the hazard rate, the probability that a firm will adopt the COLI shelter at time t , given that it has not previously adopted the COLI shelter. Although event history analysis often utilizes continuous-time hazard-rate models, the data in this study are better suited to a discrete-time model because both the time of adoption and the related explanatory variables are observed at annual intervals only (Allison 1984).²⁸ I construct a discrete-time event history model in which each firm contributes an observation for each year it is "at risk" of adopting a COLI

²⁸ Discrete-time hazard models have been used to examine the diffusion of a variety of corporate practices (e.g., Fligstein 1985; Mezas 1990; Rao and Sivakumar 1999) as well as the adoption pattern of various state policies (e.g., Berry and

shelter. Court documents indicate that the COLI shelter was developed in response to TRA86, and I identify at least one firm that appears to adopt COLI in 1986. Therefore, I assume that firms are “at-risk” if they have not adopted the COLI shelter and t is greater than 1985 and estimate a logistic regression of the following form:

$$\log(P(t)/[1 - P(t)]) = \sum a(t) + \sum b_i x_i(t) + \sum c_j x_j, \quad (4)$$

where $P(t)$ is the probability of adopting the COLI shelter, b_i is the set of coefficients for explanatory variables $x_i(t)$ that do change with time, c_j is the set of coefficients for explanatory variables x_j that do not change over time, and $a(t)$ represents separate constants for each period t . Specifically, I estimate nested models derived from the following logistic regression:

$$\begin{aligned} \text{Log}(P(t)/[1 - P(t)]) = & \beta_0 + \beta_1 \text{FOREIGN}_{i,t-1} + \beta_2 \text{RD}_{i,t-1} + \beta_3 \text{INDRANK}_{i,t} + \beta_4 \text{SGRWTH}_{i,t} \\ & + \beta_5 \text{ETR}_{i,t-1} + \beta_6 - \beta_{14} \text{yr indicators} + \beta_{15} \text{BODLINK2}_{i,t} \\ & + \beta_{16} \text{AUDITLINK2}_{i,t} + \beta_{17} \text{INDLINK2}_{i,t} + \beta_{18} \text{BEALINK2}_{i,t} \\ & + \beta_{19} \text{DISTANCE}_{i,t} + \varepsilon_{i,t}. \end{aligned} \quad (5)$$

The discrete hazard model specified in Equation (5) accommodates time-varying covariates by splitting the history of each sample firm into one-year records or spells, with all spells except the year of adoption coded as right-censored.²⁹

Duration models commonly assume that the baseline hazard rate is time-invariant and that the rate of adoption varies with firm characteristics, but is otherwise constant across time. However, as Figure 1 indicates, COLI adoptions appear to be related to firms' expectations about legislative measures to restrict the preferential tax treatment afforded to life insurance. Therefore, I include year indicators to determine how the hazard function changes over time. If the threat of potential adverse tax law changes prompted firms to complete COLI programs in the hopes of a grandfathering provision, then the rate of COLI adoption should be higher in 1989 and 1990 relative to other years. If Congress' failure to pass legislation curbing COLI benefits in 1991 and 1992 reduced the uncertainty surrounding COLI, then the rate of COLI adoption should be higher in 1993 relative to other years.

Independent Variables

In the cross-sectional model, I test H1 using a broad measure of board interlock influence, *BODLINK1*, that does not vary across time. In the event history model, I employ a more restrictive measure of board interlock influence that accounts for the timing of board interlocks and COLI adoptions. *BODLINK2* is equal to the number of board interlocks the sample firm has in year t with firms that have adopted COLI prior to year t . Consistent with H1, I expect a positive coefficient on *BODLINK2*.

The cross-sectional model also employs dichotomous variables to test H2, H3, and H4a. In the event history model, I use three alternative measures to test these hypotheses: *AUDITLINK2*, *INDLINK2*, and *BEALINK2*. *AUDITLINK2* is equal to the number of prior COLI adopters audited by the firm's auditor in year t , scaled by the total number of Compustat firms audited by the firm's

Berry 1992, 1994; Mintrom and Vergari 1998). See Box-Steffensmeier and Jones (1997) for an in-depth discussion of the benefit of using discrete-time models for certain data and Omer and Shelley (2004) for an example of the use of a discrete-time specification in the accounting literature.

²⁹ COLI adopters contribute firm years in proportion to the timing of adoption; a firm that adopts in 1986 contributes one firm year, whereas a firm that adopts in 1995 contributes ten firm years, nine coded as 0 and one coded as 1. Non-adopters contribute ten firm years, the length of the window, all coded as 0.

auditor in year t . If COLI spreads via shared auditors (H2), then the rate of COLI adoption should be higher for firms whose auditors have prior contact with COLI, resulting in a positive coefficient for *AUDITLINK2*.

INDLINK2 is equal to the number of prior COLI adopters in a firm's industry in year t , scaled by the total number of Compustat firms in the firm's industry in year t . If firms imitate their structurally equivalent peers captured by shared industry membership (H3), then the rate of COLI adoption should be higher for firms in industries with a greater frequency of prior COLI adopters, resulting in a positive coefficient for *INDLINK2*. Similarly, *BEALINK2* equals the number of prior COLI adopters in the firm's BEA region in year t , scaled by the total number of Compustat firms in the firm's BEA region in year t . If firms imitate their geographically proximate peers (H4a), then the coefficient on *BEALINK2* will be positive.

As in the cross-sectional model, I include *DISTANCE*, an ordinal variable that proxies for the geographic distance between a firm's headquarters and the center of early COLI activity. I expect firms headquartered within or close to the BEA's Mid-East region to adopt earlier relative to firms headquartered farther away (H4b); therefore, I expect the rate of COLI adoption to be negatively associated with *DISTANCE*.

Discrete-Time Hazard Logit Results

Table 6 reports results from estimating the discrete-time hazard model in Equation (5).³⁰ All models have reasonable explanatory power, with pseudo- R^2 values ranging from 15.62 percent to 19.29 percent. The coefficients on the year indicators for 1989, 1990, 1993, and 1994 are positive and significant, indicating that the hazard rate, the likelihood of adopting COLI in year t given that the firm has not adopted COLI prior to year t , is significantly greater in those years relative to the base year, 1986. This result is consistent with firms timing their adoptions in response to concerns over legislative restrictions. As in the cross-sectional model, *FOREIGN_{t-1}* is the only control variable that is significant in the predicted direction across all models.

Consistent with H1, the coefficient on *BODLINK2* is positive and significant (1.67; $z = 3.72$; $p < 0.01$); the conditional likelihood of adopting COLI in year t is greater for firms with board ties to prior COLI adopters. In marginal effect terms, given that the firm has not already adopted COLI, in 1993, a firm with one board tie to a prior COLI adopter is 2.4 times as likely to adopt COLI relative to a firm with no board ties (increase in probability of adopting from 11.97 percent to 40.74 percent).³¹

Contrary to H2 and H3, but similar to the results in Table 5, the coefficients on *AUDITLINK2* and *INDLINK2* in the discrete-time hazard model are not significant. While *DISTANCE* is significant in the cross-sectional models using a one-to-one control sample and both *DISTANCE* and *BEALINK1* are significant in the cross-sectional model using a one-to-many control sample, neither *DISTANCE* nor *BEALINK2* is significant in the discrete-time hazard model. One interpretation of these mixed results is that *DISTANCE* affects the likelihood of adopting COLI but, contrary to expectation, *DISTANCE* is not related to the timing of adoptions.³²

³⁰ Three COLI firms do not have data available to calculate the explanatory variables in all years from 1986 to 1996; thus, the sample used to estimate Equation (5) consists of 629 firm-year observations representing 38 COLI firms and 38 matched control firms.

³¹ The marginal effect is calculated by setting the indicator variable for 1993, *FOREIGN_{t-1}* and *RD_{t-1}* equal to 1, all other year indicator variables equal to 0, and all continuous and ordinal variables equal to their median values.

³² In untabulated results using a Tobit specification, *DISTANCE* is negatively and significantly related to *DURATION*, which equals the total number of years from the time the firm started using COLI until 1996 for COLI firms, and 0 for control firms (coeff = -1.88; $t = -1.99$; one-tailed $p = 0.03$). The Tobit results are consistent with H4b, firms headquartered farther away from the Mid-East BEA region adopt later.

TABLE 6
Discrete-Time Hazard Model

$$\text{Log}(P(t)/[1 - P(t)]) = \beta_0 + \beta_1 \text{FOREIGN}_{i,t-1} + \beta_2 \text{RD}_{i,t-1} + \beta_3 \text{INDRANK}_{i,t} + \beta_4 \text{SGRWTH}_{i,t} + \beta_5 \text{ETR}_{i,t-1} + \beta_6 - \beta_1 \Delta \text{yr indicators}$$

$$+ \beta_7 \text{BODLINK2}_{i,t} + \beta_8 \text{AUDITLINK2}_{i,t} + \beta_9 \text{INDLINK2}_{i,t} + \beta_{10} \text{BEALINK2}_{i,t} + \beta_{11} \text{DISTANCE}_{i,t} + \varepsilon_{i,t}$$

Variable	Pred. Sign	Base		H1: BOD Interlock		H2: Auditor		H3: Industry	
		Coeff	z-stat	Coeff	z-stat	Coeff	z-stat	Coeff	z-stat
Intercept		-6.247	-3.07***	-5.991	-2.92***	-6.419	-3.14***	-6.258	-3.09***
FOREIGN _{t-1}	-	-1.204	-2.40***	-1.084	-2.04***	-1.155	-2.29***	-1.202	-2.39***
RD _{t-1}	-	0.766	1.92	0.730	1.71	0.770	1.91	0.745	1.69
INDRANK	+	0.176	1.06	0.134	0.82	0.191	1.15	0.178	1.08
SGRWTH	+	-2.044	-1.91	-1.304	-1.22	-1.975	-1.81	-2.041	-1.90
ETR _{t-1}	?	1.036	0.87	1.217	1.05	1.091	0.90	1.037	0.87
D-1989		1.980	1.71*	1.812	1.55	2.045	1.75*	1.974	1.70*
D-1990		2.131	1.89*	1.916	1.70*	2.263	1.97**	2.115	1.85*
D-1993		2.962	2.70***	2.796	2.54**	3.236	2.79***	2.928	2.57**
D-1994		3.046	2.73***	2.665	2.39**	3.454	2.76***	2.991	2.56**
BODLINK2	+			1.670	3.72***				
AUDITLINK2	+					-0.669	-0.81		
INDLINK2	+							0.058	0.19
BEALINK2	+								
DISTANCE	-								
Goodness of Fit Statistics									
Pseudo-R ²		15.62%		19.17%		15.86%		15.62%	
Chi-square test		35.60***		51.36***		36.50***		37.59***	
Area under ROC curve		0.80		0.81		0.80		0.80	

(continued on next page)

Variable	Pred. Sign	H4a: BEA Region		H4b: Distance		H1 and H4b	
		Coeff	z-stat	Coeff	z-stat	Coeff	z-stat
Intercept		-6.341	-3.10***	-5.852	-2.75***	-5.761	-2.67***
$FOREIGN_{t-1}$	-	-1.200	-2.44***	-1.186	-2.36***	-1.072	-2.00***
RD_{t-1}	-	0.711	1.74	0.711	1.76	0.695	1.59
$INDRANK$	+	0.186	1.11	0.195	1.19	0.145	0.90
$SGRWTH$	+	-2.096	-1.93	-2.014	-1.84	-1.297	-1.20
ETR_{t-1}	?	1.112	0.99	0.858	0.71	1.115	0.94
$D-1989$		1.914	1.64*	1.931	1.67*	1.791	1.53
$D-1990$		1.981	1.75*	2.094	1.86*	1.902	1.69*
$D-1993$		2.691	2.37**	2.919	2.66***	2.776	2.52**
$D-1994$		2.694	2.30**	3.027	2.73***	2.664	2.40**
$BODLINK2$	+					1.621	3.53***
$AUDITLINK2$	+						
$INDLINK2$	+						
$BEALINK2$	+	0.583	1.08				
$DISTANCE$	-			-0.224	-1.10	-0.129	-0.60
Goodness of Fit Statistics							
Pseudo-R ²		16.02%		16.02%		19.29%	
Chi-square test		38.05**		37.66***		49.86***	
Area under ROC curve		0.80		0.80		0.81	

*, **, *** Indicate significance at the 0.10, 0.05, and 0.01 levels in two-tailed tests (one-tailed when predicted).
The sample includes 629 firm-year observations representing 38 COLI firms and 38 control firms from a sample matched on *ROA* and *lnEMP* in year $t-1$ (the year before adoption).
Standard errors are calculated after clustering on firm.

Variable Definitions:

$FOREIGN_{t-1}$ = indicator variable equal to 1 if the absolute value of foreign pre-tax income (PIFO) divided by the absolute value of total pre-tax income (PTI) is greater than 10 percent;
 RD_{t-1} = indicator variable equal to 1 if R&D expense (XRD) is positive;
 $INDRANK$ = decile-ranking resulting from ranking all Compustat firms by year, industry, and total assets (AT);
 $SGRWTH$ = sales (SALE) in year t minus sales in year $t-1$, divided by sales in year $t-1$;
 ETR_{t-1} = lagged value of total income taxes (TXT) divided by pre-tax book income (PTI);
 $BODLINK2$ = number of board interlocks the sample firm has in year t with firms that have adopted COLI prior to year t ;
 $AUDITLINK2$ = number of prior COLI adopters audited by the firm's auditor in year t , scaled by the total number of Compustat firms audited by the firm's auditor in year t ;
 $INDLINK2$ = number of prior COLI adopters in a firm's industry in year t , scaled by the total number of Compustat firms in the firm's industry in year t ;
 $BEALINK2$ = number of prior COLI adopters in the firm's BEA region in year t , scaled by the total number of Compustat firms in the firm's BEA region in year t ; and

DISTANCE = ordinal variable equal to 1 for firms headquartered in the Mid-East BEA region, 2 for firms headquartered in the neighboring New England, Southeast, and Great Lakes BEA regions, 3 for firms headquartered in the Plains and Southwest BEA regions, and 4 for firms headquartered in the Rocky Mountain and Far West BEA region.

VI. SUMMARY AND CONCLUSION

This study contributes to the literature on aggressive corporate tax reporting by examining how the use of a particular corporate tax shelter, corporate-owned life insurance (COLI), spreads over time and across firms. I develop hypotheses based on four social environment factors derived from the literature on innovation-diffusion and mimetic isomorphism and test whether these factors help to explain the pattern of COLI adoptions. I find that COLI adopters are connected through board interlocks. Consistent with a cohesion model of diffusion, a direct tie to a prior COLI adopter via a board interlock increases the likelihood that a firm adopts the COLI shelter. I find no evidence that firms' connections to prior COLI adopters via a common auditor are related to the spread of COLI adoption. I also examine whether imitation of structurally equivalent peers helps to explain the spread of COLI participation. However, I find that the number of prior adopters in a firm's industry is not related to the likelihood that it will adopt COLI. Finally, I find some evidence that COLI use spreads geographically.

This study specifically examines which diffusing mechanisms help explain the spread of COLI use. Not all corporate practices diffuse the same way (Davis and Greve 1997), and the importance of a particular diffusion mechanism, like board ties, can wane over time (Mizruchi et al. 2006). Although the findings presented here do not necessarily generalize to all corporate tax shelters or to the adoption of tax strategies in today's information environment, the results related to COLI highlight the importance of examining firms' decisions in light of the broader social structures in which they operate. Moreover, the theories of innovation-diffusion and institutional isomorphism explored in this study are potentially applicable to a variety of accounting practice trends.

APPENDIX

Accounting Treatment of COLI

The level of detail, if any, disclosed in the financial statement footnotes regarding COLI plans varies considerably from firm to firm.³³ Investments in life insurance are governed by FASB Technical Bulletin No. 85-4, which prescribes the use of the cash surrender value (CSV) method. Under the CSV method, a firm should record the CSV of the life insurance policy as an asset, but under the right of setoff, firms can net outstanding policy loans against the CSV of the policy, rather than record a separate liability (APB No. 10, FASB Technical Bulletin No. 88-2, and FASB Interpretation No. 39). Given the right of setoff, firms can participate in the COLI shelter with little net effect on their balance sheets.³⁴

As seen in the example below, other than through the tax effect, COLI has little income statement effect. In a leveraged COLI transaction, net COLI income (expense) equals increases in the cash surrender value (interest credited to the policy) plus any death benefits received less the premium expense and the interest expense on the policy loans (FASB Technical Bulletin No. 85-4). SFAS No. 109, *Accounting for Income Taxes*, indicates that the excess of CSV over premiums paid results in a permanent book-tax difference if the insurance policy is expected to be held until the death of the insured. Thus, COLI programs that produce material book-tax differences should be reported in the firm's tax rate reconciliation. However, firms have considerable discretion over how permanent differences are netted and aggregated into specific line items on the rate reconciliation.

³³ Material for this section is drawn from Nurnberg (2004).

³⁴ Court documents reveal that Camelot's COLI plan was set to maintain a zero net equity balance at the end of each year (*In re C.M. Holdings, Inc.*, 254 B.R. 578 (D. Del. 2000)).

	Book	Book-Tax Adjustment	Tax
Increase in policy cash surrender value	20	(20)	0
Less: Premium expense	(12)	12	0
Interest expense on policy loans	(10)	0	(10)
Net COLI income (expense)	(2)	(8)	(10)

Under IRC §72, the interest credited to the cash surrender value of the policy, \$20 in the example above, is not included in the policyholder's gross income. Because proceeds from life insurance receive preferential tax treatment, policy premiums are not deductible for tax purposes. This example shows premium expense of \$12, which reduces book income, but not taxable income. However, the interest expense paid on policy-backed loans, here \$10, is tax deductible.

The numbers chosen for the example here are arbitrary, but show an important feature of the COLI shelter. The transaction produces a loss for tax purposes in excess of any economic or financial statement loss. Indeed, even though this example produces a \$2 pre-tax book loss, at a 35 percent statutory rate, the \$2.80 tax benefit from \$8 of net nontaxable income offsets the \$2 pre-tax book loss, generating positive after-tax income of \$0.80.

REFERENCES

- Abrahamson, E., and L. Rosenkopf. 1993. Institutional and competitive bandwagons: Using mathematical modeling as a tool to explore innovation diffusion. *Academy of Management Review* 18 (3): 487–517.
- Allison, P. D. 1984. *Event History Analysis*. Beverly Hills, CA: Sage.
- Bankman, J. 1999. The new market for corporate tax shelters. *Tax Notes* 83 (12): 1775–1795.
- . 2002. Commentary: Modeling the tax shelter world. *Tax Law Review* 55: 455–464.
- . 2004. An academic's view of the tax shelter battle. In *The Crisis in Tax Administration*, edited by H. J. Aaron, and J. Slemrod. Washington, D.C.: Brookings Institution.
- Barth, M. E., W. H. Beaver, and W. R. Landsman. 1998. Relative valuation roles of equity book value and net income as a function of financial health. *Journal of Accounting and Economics* 25 (1): 1–34.
- Berry, F. S., and W. D. Berry. 1990. State lottery adoptions as policy innovations: An event history analysis. *The American Political Science Review* 84: 395–415.
- , and ———. 1992. Tax innovation in the states: Capitalizing on political opportunity. *American Journal of Political Science* 36: 715–742.
- , and ———. 1994. The politics of tax increases in the states. *American Journal of Political Science* 38: 855–859.
- Bizjak, J., M. Lemmon, and R. Whitby. 2009. Option backdating and board interlocks. *Review of Financial Studies* 22 (11): 4821–4847.
- Box-Steffensmeier, J., and B. Jones. 1997. Time is of the essence: Event history models in political science. *American Journal of Political Science* 41 (4): 1414–1461.
- Burns, L. R., and D. R. Wholey. 1993. Adoption and abandonment of matrix management programs: Effects of organizational characteristics and interorganizational networks. *Academy of Management Journal* 36 (1): 106–138.
- Burt, R. S. 1987. Social contagion and innovation: Cohesion versus structural equivalence. *American Journal of Sociology* 92: 1287–1335.
- Cyert, R., and J. G. March. 1963. *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ: Prentice Hall.
- Davis, G. F. 1991. Agents without principles? The spread of the poison pill through the intercorporate network. *Administrative Science Quarterly* 36: 605–633.
- , and H. R. Greve. 1997. Corporate elite networks and governance changes in the 1980s. *American Journal of Sociology* 103: 1–37.
- Desai, M., and D. Dharmapala. 2006. Corporate tax avoidance and high-powered incentives. *Journal of Financial Economics* 79: 145–179.

- DiMaggio, P. J., and W. W. Powell. 1983. The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review* 48: 147–160.
- Eustice, J. 2002. Abusive tax shelters: Old “brine” in new bottles. *Tax Law Review* 55: 135–174.
- Fligstein, N. 1985. The spread of the multidivisional form among large firms, 1919–1979. *American Sociological Review* 50: 377–391.
- . 1990. *The Transformation of Corporate Control*. Cambridge, MA: Harvard University Press.
- Frank, M., L. Lynch, and S. Rego. 2009. Are financial and tax reporting aggressiveness reflective of broader corporate policies. *The Accounting Review* 84 (2): 467–498.
- Galaskiewicz, J., and R. Burt. 1991. Interorganization contagion in corporate philanthropy. *Administrative Science Quarterly* 36 (1): 88–105.
- Gergen, M. 2002. The logic of deterrence: corporate tax shelters. *Tax Law Review* 55: 255–286.
- Graham, J. R., and A. L. Tucker. 2006. Tax shelters and corporate debt policy. *Journal of Financial Economics* 81: 563–594.
- Gupta, S., and K. Newberry. 1997. Determinants of the variability in corporate effective tax rates: Evidence from longitudinal data. *Journal of Accounting and Public Policy* 16 (1): 1–34.
- Han, S.-K. 1994. Mimetic isomorphism and its effect on the audit services market. *Social Forces* 73 (2): 637–664.
- Haunschild, P. R. 1993. Interorganizational imitation: The impact of interlocks on corporate acquisition activity. *Administrative Science Quarterly* 38 (4): 564–592.
- . 1994. How much is that company worth? Interorganizational relationships, uncertainty and acquisition premiums. *Administrative Science Quarterly* 39: 391–412.
- , and A. S. Miner. 1997. Modes of interorganizational imitation: The effects of outcome salience and uncertainty. *Administrative Science Quarterly* 42 (3): 472–500.
- Haveman, H. 1993. Follow the leader: Mimetic isomorphism and entry into new markets. *Administrative Science Quarterly* 38 (4): 593–627.
- Lisowsky, P. 2010. Seeking shelter: Empirically modeling tax shelters using financial statement information. *The Accounting Review* 85 (5): 1693–1720.
- Mahajan, V., and R. Peterson. 1985. *Models for Innovation Diffusion*. Beverly Hills, CA: Sage Publications.
- March, J. G. 1981. Decisions in organizations and theories of choice. In *Perspectives on Organization Design and Behavior*, edited by Van de Ven, A., and W. F. Joyce. New York, NY: Wiley.
- Mezias, S. 1990. An institutional model of organizational practice: Financial reporting at the Fortune 200. *Administrative Science Quarterly* 35 (3): 431–457.
- Mintrom, M., and S. Vergari. 1998. Policy networks and innovation diffusion: The case of state education reforms. *The Journal of Politics* 60 (1): 126–148.
- Mizruchi, M. S. 1992. *The Structure of Corporate Political Action*. Cambridge, MA: Harvard University Press.
- , L. B. Stearns, and C. Marquis. 2006. The conditional nature of embeddedness: A study of borrowing by large U.S. firms, 1973–1994. *American Sociological Review* 71 (2): 310–333.
- Novack, J., and L. Saunders. 1998. The hustling of X-rated shelters. *Forbes* (December): 19.
- Nurnberg, H. 2004. Accounting for company-owned life insurance. *Accounting Horizons* 18 (2): 109–126.
- Omer, T. C., and M. K. Shelley. 2004. Competitive, political and economic factors influencing state tax policy changes. *The Journal of the American Taxation Association* 26 (Supp.): 103–126.
- Palmer, D. A., P. D. Jennings, and X. Zhou. 1993. Late adoption of the multidivisional form by large U.S. corporations: Institutional, political, and economic accounts. *Administrative Science Quarterly* 38: 100–131.
- Paull, L. 2000. JCT Testimony at Finance Committee hearing on Code’s penalty and interest provisions. *Tax Notes Today* (March 9).
- Pfeffer, J. 1981. Management as symbolic action: The creation and maintenance of organizational paradigms. *Research in Organizational Behavior* 3: 1–52.
- Rao, H., and K. S. vakumar. 1999. Institutional sources of boundary-spanning structures: The establishment of investor relations departments in the Fortune 500 industrials. *Organization Science* 10 (1): 27–42.
- Rogers, E. 2003. *Diffusion of Innovations*. 5th edition. New York, NY: Free Press.
- Strang, D., and S. A. Soule. 1998. Diffusion in organizations and social movements: From hybrid corn to

- poison pills. *Annual Review of Sociology* 24: 265–290.
- Stuart, T. E., and S. Yim. 2010. Board interlocks and the propensity to be targeted in private equity transactions. *Journal of Financial Economics* 97: 174–189.
- U.S. Department of the Treasury. 1990. Report to Congress on the taxation of life insurance company products. Washington, D.C.: U.S. Government Printing Office.
- . 1999. *The Problem of Corporate Tax Shelters: Discussion, Analysis, and Legislative Proposals*. Washington, D.C.: U.S. Government Printing Office.
- U.S. Senate Permanent Subcommittee on Investigations. 2005. *The Role of Professional Firms in the U.S. Tax Shelter Industry*. S. Rep. 109-54. Washington, D.C.: U.S. Government Printing Office.
- Weisbach, D. 2002. Ten truths about corporate tax shelters. *Tax Law Review* 55: 215–253.
- Wilson, R. 2009. An examination of corporate tax shelter participants. *The Accounting Review* 84 (3): 969–999.

Voluntary Nonfinancial Disclosure and the Cost of Equity Capital: The Initiation of Corporate Social Responsibility Reporting

Dan S. Dhaliwal

The University of Arizona and Korea University

Oliver Zhen Li

The University of Arizona

Albert Tsang

Yong George Yang

The Chinese University of Hong Kong

ABSTRACT: We examine a potential benefit associated with the initiation of voluntary disclosure of corporate social responsibility (CSR) activities: a reduction in firms' cost of equity capital. We find that firms with a high cost of equity capital in the previous year tend to initiate disclosure of CSR activities in the current year and that initiating firms with superior social responsibility performance enjoy a subsequent reduction in the cost of equity capital. Further, initiating firms with superior social responsibility performance attract dedicated institutional investors and analyst coverage. Moreover, these analysts achieve lower absolute forecast errors and dispersion. Finally, we find that firms exploit the benefit of a lower cost of equity capital associated with the initiation of CSR disclosure. Initiating firms are more likely than non-initiating firms to raise equity capital following the initiations; among firms raising equity capital, initiating firms raise a significantly larger amount than do non-initiating firms.

Keywords: *corporate social responsibility; cost of capital; voluntary disclosure.*

Data Availability: *The data are publicly available from the sources identified in the paper.*

We appreciate the valuable comments from Ying Cao, Zhaoyang Gu, Michel Magnan, Morton Pincus, Suresh Radhakrishnan, T. J. Wong, and seminar participants at the City University of Hong Kong, Singapore Management University, Tsinghua University and 2009 AAA Annual Meeting, and especially the detailed constructive suggestions from Steven Kachelmeier (editor) and two anonymous referees. We thank Brian Bushee for generously sharing his classifications of institutional investors. We also thank CorporateRegister.com and KLD Research & Analytics, Inc. for providing the corporate social responsibility data, and I/B/E/S for the analyst following data. We are grateful to Xuemei Hao and Guangqing Li for their excellent research assistance. The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong SAR, China (Project No. CUHK451409).

Editor's note: Accepted by Steven Kachelmeier.

Submitted: April 2009

Accepted: May 2010

Published Online: January 2011

I. INTRODUCTION

The last 15 years have witnessed a steadily increasing emphasis on socially responsible corporate activities around the world. While third parties, such as KLD Research and Analytics, Inc. (KLD), often track and rate the corporate social responsibility (CSR) performance of large firms, firms have also become increasingly willing to voluntarily issue standalone CSR reports in recent years.¹ According to CorporateRegister.com, a private company that specializes in tracking CSR reports, few standalone CSR reports were issued in the United States before the mid-1990s. However, since then, increasingly more U.S. firms have committed to making this type of disclosure. In 2007 alone, large firms issued about 300 CSR reports. Although CSR disclosing firms represent only a small fraction of the population of U.S. publicly listed firms, their aggregate market value constituted over 10 percent of the total U.S. market capitalization in 2007.² The rapid increase in CSR reporting naturally raises questions among researchers: What are the rationales behind this type of voluntary disclosure? What benefits do firms gain by spending resources on compiling and publishing these standalone reports, especially given that CSR performance ratings are often available to investors through third parties?

A number of factors potentially provide answers to these questions, such as the growing influence of global enterprises, the intensified scrutiny of corporate impact on the society and the economy as a result of a loss of trust after a series of corporate scandals around 2001, and the recent rapid growth in ethical/socially responsible investment in the United States and around the world.³ Anecdotal evidence also indicates that firms' reputations and long-term sales can suffer because of poor CSR performance. For instance, Nike struggled for years and invested a great amount of financial resources and effort to regain its reputation after the 1997 child labor scandal.⁴

We examine one factor, namely, a reduction in firms' cost of equity capital, that potentially provides an explanation for the increasing trend in CSR disclosure. Among various potential factors influencing CSR disclosure decisions, we focus on the cost of equity capital because it plays a critical role in a firm's financing and general operations decisions. Also, corporate executives appear to believe that voluntarily communicating information can reduce their firms' cost of capital (Graham et al. 2005). Further, there is a longstanding interest among academics in the relation between disclosure and the cost of capital (Diamond and Verrecchia 1991; Botosan 1997; Leuz and Verrecchia 2000; Botosan and Plumlee 2002).

To determine whether and how CSR disclosure is related to firms' cost of equity capital, we employ a sample of firms that intersect two CSR data sources: (1) a comprehensive list of firms releasing electronic or hard-copy standalone CSR reports since 1993, collected from various sources on the Internet; and (2) the KLD STATS database that provides detailed CSR performance ratings for individual firms. Our analyses provide four important insights. First, firms with a high cost of equity capital in the previous year are significantly more likely than others to initiate standalone CSR disclosures. Second, the cost of equity capital decreases for CSR-initiating firms with superior CSR performance. Third, CSR-initiating firms with superior CSR performance attract dedicated institutional investors and analyst coverage. Moreover, these analysts have more

¹ Consistent with McWilliams and Siegel (2001), among others, we define CSR as instances where the company goes beyond compliance and voluntarily engages in actions that appear to advance social causes, including committing to environmental and human rights protection, providing community support, and so forth. In practice and academic research, CSR is often used interchangeably with "sustainability." We also follow this convention in the paper.

² This figure is based on the mean market cap of \$14.47 billion for firms as represented in Table 2 and the total U.S. market cap of around \$15.35 trillion on May 23, 2007.

³ For example, according to the Social Investment Forum (2007), from 1995 to 2005, assets invested in socially responsible investment grew from \$639 billion to \$2.29 trillion, and accounted for approximately 11 percent of the total assets managed by professional managers.

⁴ See <http://www.bandt.com.au/news/25/0c00d225.asp>.

accurate forecasts and lower forecast dispersion. Finally, corroborating the result on the relation between CSR disclosure and the cost of equity capital, CSR-initiating firms are significantly more likely than non-initiating firms to conduct seasoned equity offerings (SEOs) in the two years following these initiations and among firms conducting SEOs, CSR-initiating firms raise a significantly larger amount of capital than do non-initiating firms. Overall, our evidence is consistent with our predictions that a potential reduction in the cost of equity capital motivates firms to publish standalone CSR reports and that CSR disclosure by firms with superior CSR performance leads to a lower cost of equity capital.

This study is the first to investigate the impact of standalone voluntary disclosure of general CSR issues on the cost of equity capital. We contribute to the literature by extending the traditional research on voluntary disclosure beyond the narrow focus of financial disclosure. The extant finance and accounting literatures on voluntary disclosure focus primarily on management forecasts or conference calls that are short-term-oriented.⁵ In contrast, CSR disclosure, which is broad in scope, is related to a firm's long-term development strategies and performance sustainability. Our results provide evidence on the rationales behind and the consequences of the recent trend in voluntary CSR disclosure.

Our study is related to, but differs from, the work of Plumlee et al. (2008) and Richardson and Welker (2001). Plumlee et al. (2008) examine the impact of voluntary environmental disclosure quality on firm value. We examine a broader concept of CSR, which includes environmental protection, community development, corporate governance practices, employee relations, diversity practices, human rights, and product quality. In addition, we use a measure of CSR that is different from Plumlee et al. (2008), who use a self-constructed index to measure firms' environmental disclosure quality. We use a proxy that indicates whether firms publish CSR reports. Also, the information examined by Plumlee et al. (2008) comes from corporate environmental reports as well as annual reports and 10-Ks, which reflect both voluntary and mandatory disclosures. The standalone CSR reports we examine are voluntary.

Our study also differs from Richardson and Welker (2001), who examine the relation between the cost of equity capital and social as well as financial disclosure. First, we study U.S. firms, whereas they examine Canadian firms. The United States and Canada differ considerably in institutions related to information disclosure, with the United States having more stringent regulations than Canada (Richardson and Welker 2001). If more stringent regulations and the associated higher level of litigation risk translate into a generally higher level of disclosure credibility, then we can observe different relations between disclosure and the cost of equity capital in these two countries. In addition, the CSR measure used by Richardson and Welker (2001) is based on annual reports, whereas we focus on standalone CSR disclosures. These two forms of disclosure differ in depth and breadth of CSR coverage.

Methodologically, we differ from Plumlee et al. (2008) and Richardson and Welker (2001) by employing a lead-lag approach enhanced with two-stage regressions in sensitivity analysis to deal with endogeneity and self-selection issues and by exploring the underlying channels, such as institutional ownership and analyst coverage, through which CSR disclosure affects the cost of equity capital. In sum, we contribute to the literature by complementing and extending Plumlee et al. (2008) and Richardson and Welker (2001).

Section II develops our hypotheses. Section III describes our sample and methodology. Section IV presents empirical evidence on the relation between CSR disclosure and the cost of equity capital. Section V summarizes and concludes.

⁵ One of the few exceptions is Dietrich et al. (2001), who investigate the effect of the supplemental disclosure of forward-looking information on security prices.

II. RELATED RESEARCH AND HYPOTHESIS DEVELOPMENT

Most prior research on the relation between disclosure and the cost of capital focuses on financial disclosure (Core 2001; Healy and Palepu 2001; Leuz and Wysocki 2008). The consensus appears to be that a negative relation exists between the quality of financial disclosure and the cost of capital. Greater disclosure increases investors' awareness of a firm's existence and enlarges its investor base, which improves risk-sharing and reduces the cost of capital (Merton 1987). In addition, higher quality or more precise firm-specific disclosures decrease the covariance of a firm's cash flow with the cash flows of other firms (Hughes et al. 2007; Lambert et al. 2007), which essentially reduces the betas of individual firms and, hence, the cost of equity capital. Similarly, greater disclosure can lead to reduced information asymmetry among investors or between managers and investors. When the level of disclosure is inadequate and some investors are perceived to be better informed than others, informationally disadvantaged investors price-protect themselves and become less willing to trade. The resultant illiquidity increases the bid-ask spread and transaction costs (Verrecchia 2001), which leads to a higher required rate of return or cost of equity capital (Amihud and Mendelson 1986).

These mechanisms likely apply to both financial and nonfinancial disclosure, as long as the information concerned is value-relevant. Indeed, a fair amount of research suggests that CSR information is value-relevant (Margolis and Walsh 2001; Orlitzky et al. 2003; Al-Tuwaijri et al. 2004). Of course, CSR practices can affect firms' financial performance and value through channels other than those related to financial disclosure. For instance, voluntary socially responsible behavior can help firms avoid government regulation and, therefore, reduce compliance costs. In addition, socially responsible firms appeal to consumers who care about the corresponding social issues, which leads to superior sales and financial performance (Lev et al. 2010). Socially aware investors are willing to pay a premium for the securities of socially responsible firms (Anderson and Frankel 1980; Richardson and Welker 2001). Perhaps more important, some CSR projects have direct implications for positive cash flow even in the near future. For example, practices related to protecting the environment and improving employee welfare can reduce potential litigation and pollution cleaning costs, boost employee morale. And, thereby, production efficiency. These arguments highlight the importance of CSR disclosure in reducing information asymmetry and uncertainty related to factors affecting firm value (Rodriguez et al. 2006), which in turn reduces the cost of equity capital.

Nevertheless, a straightforward generalization of the cost of capital effect from financial disclosure to nonfinancial CSR disclosure is not always obvious. Standalone CSR reports are currently subject to very limited regulatory guidance. There is a common concern about the usefulness of this type of disclosure because of noncomparability and potential credibility issues and opportunistic behaviors of firms (Ingram and Frazier 1980; Hobson and Kachelmeier 2005).⁶ In the end, whether *voluntary* CSR disclosure reduces a firm's cost of equity capital is an empirical question.

It is important to note that CSR performance ratings of large firms are often available to investors through third parties. These ratings could be directly associated with the cost of equity capital of these firms. However, ratings alone are unlikely to provide sufficient information for investors to assess firms' overall CSR performance. Detailed CSR disclosures potentially provide additional information necessary for investors to assimilate these summary ratings.⁷ Further, voluntarily disclosing CSR activities demonstrates firms' confidence in their CSR performance, which

⁶ Although some accounting and consulting firms provide voluntary assurance service (Simmnett et al. 2009), there is not yet a government standard that regulates this service, and the assurance industry is still in its infancy.

⁷ An obvious analogy is the usefulness of footnote disclosures and management discussions in supplementing financial statements.

sends a positive signal to investors, or, in the case of poor CSR performance, allows firms to offer explanations. Therefore, CSR disclosures contain information beyond that contained in CSR performance ratings.

Some firms also disclose information on CSR activities in their annual reports or filings with the SEC. However, a firm's voluntary compilation and publication of standalone CSR reports demonstrates its special effort and commitment to improving transparency regarding long-term performance and risk management. More importantly, compared with the CSR information provided in annual reports or 10-Ks, standalone CSR reports are more comprehensive and contain significantly more details.⁸ Therefore, standalone CSR reports likely provide incrementally useful information for investors to evaluate firms' long-term sustainability. Focusing on standalone CSR reports can thus improve the power of our tests and shed light on this new form of voluntary nonfinancial disclosure.

Our first hypothesis predicts that a possible reduction in the cost of equity capital provides an incentive for firms to publish CSR reports. Frankel et al. (1995) find that firms increase their level of voluntary disclosure to raise capital in the future at a lower cost, which suggests that firms with a relatively higher cost of capital likely have a greater incentive to enhance disclosure. Lending support to the cost of capital incentive for disclosure, Sletten (2008) finds that stock price declines, which imply an increase in firms' cost of equity capital, induce managers to disclose more information.⁹

Of course, endogeneity and self-selection issues can arise if we examine a contemporaneous relation between CSR disclosure and the cost of equity capital. On the one hand, if CSR disclosure is motivated by a firm's desire to reduce its high cost of equity capital, then we should find a positive relation between CSR disclosure and the cost of the equity capital. On the other hand, if CSR disclosure leads to a lower cost of equity capital, then we should find a negative relation between CSR disclosure and the cost of equity capital. Therefore, the contemporaneous relation between CSR disclosure and the cost of equity capital could be ambiguous. To address the potential endogeneity and self-selection issues related to CSR disclosure and the cost of equity capital, we employ a lead-lag approach in our main analyses and state our first hypothesis below:

H1: The likelihood that a firm will disclose its corporate social responsibility activities is positively associated with its cost of equity capital in the previous year.

If CSR disclosure provides information that is incremental to information provided in third-party CSR performance ratings or other information dissemination channels such as annual reports or 10-Ks, then the preceding discussion suggests that CSR disclosure should lead to a lower cost of equity capital. This logic suggests the following hypothesis:

H2: Corporate social responsibility disclosure is associated with a subsequently lower cost of equity capital.

Support for H1 and H2 would provide justification for the rationales behind and the consequences of CSR disclosure. We also test a corollary of H1 and H2 by examining whether disclos-

⁸ In untabulated analyses and relying on manual data collection, we compare CSR-related content in the first-time standalone CSR reports and annual reports (or 10-Ks in the absence of annual reports) of 50 firms out of our final sample of 213 firms. We find that, on average, standalone CSR reports are significantly longer (28.3 pages versus 1.5 pages) and cover significantly more CSR issues (6.4 issues versus 1.5 issues) compared to annual reports or 10-Ks. The inference of the above comparison is also supported by a comprehensive survey conducted by KPMG (2008), which finds that among the largest 100 U.S. firms, only about 1 percent of them adequately integrate CSR reports into their annual reports.

⁹ However, the result documented by Sletten (2008) could be attributable to either the numerator (cash flow) effect or the denominator (cost of capital) effect, or both.

ing firms seek external financing after CSR disclosures. If CSR disclosure is motivated by firms' desire to reduce the cost of equity capital, then these firms will be more likely than non-disclosing firms to raise equity capital after their CSR disclosures to exploit the reduction in their cost of equity capital, and they will also strive to raise a larger amount. While we formulate our predictions based on CSR disclosure, in the empirical analysis we focus on CSR-disclosure-initiating firms since initial reports likely contain more information than mundane continuing reports.

III. SAMPLE AND METHODOLOGY

Sample Description

CSR disclosure policies can be sticky across years. Therefore, we focus on first-time standalone CSR reports. We collect standalone CSR reports issued by U.S. firms from various sources, including (1) Corporate Social Responsibility Newswire, (2) CorporateRegister.com, (3) Internet searches, and (4) company websites. The first two sources are the two leading organizations collecting and disseminating news and information related to CSR. We verify our CSR reporting sample by checking whether we can find their actual standalone CSR reports.¹⁰

In our main analyses, we control for the relative social responsibility performance of sample firms, as proxied for by the KLD social performance rating scores. Our final sample comprises firms that are in both the KLD STATS and Compustat databases. KLD evaluates CSR performance for all covered firms along a variety of dimensions, regardless of whether they release standalone reports.¹¹ Starting from 1991, KLD STATS rated approximately 650 companies every year, comprising mainly all firms in the S&P 500 and Domini 400 Social SM Index. During 2001 to 2002, KLD expanded its coverage to include the largest 1,000 U.S. companies by market capitalization. Since 2003, it has covered the largest 3,000 U.S. companies based on market capitalization.

Table 1, Panel A shows the industry distribution, based on Barth et al.'s (1998) industry classifications, of CSR reports and disclosing firms. During the 1993–2007 period, 294 firms issued a total of 1,190 standalone CSR reports.¹² The Utilities industry has the largest proportion (30.4 percent) of firms publishing CSR reports, while the Services and Insurance/Real Estate industries have the lowest proportion of disclosing firms (2.15 percent and 0.40 percent, respectively). Consistent with the broad scope of CSR disclosure, many non-pollution-prone industries including the Food and Retail industries also actively disclose their social performance. After eliminating 81 firms because of missing data, our final sample contains 213 disclosing firms. The Utilities industry constitutes the largest proportion of the final sample (13.4 percent). Table 1, Panel B presents the distribution by year of CSR reports and disclosing firms. Overall, there is a steadily increasing trend in the number of CSR reports over time from 8 in 1993 to 184 in 2007. The average report length nearly doubles from about 20 pages in the early 1990s to more than 40 pages in the most recent years. On average, a CSR report has 36 pages.¹³

¹⁰ It is tempting to examine the information content of CSR disclosures. However, this test is hampered by the lack of information on the exact reporting dates of the reports. Nevertheless, we conduct an event study based on the reporting months of the reports. We find that (1) during the CSR reporting month, there is no difference in raw and market-adjusted returns between high and low CSR performance firms; (2) during the three-month period following the CSR reporting month, high CSR performance firms appear to do slightly better than low CSR performance firms, based on market-adjusted returns; and (3) there is no difference in returns between CSR reporting months and non-CSR reporting months.

¹¹ The Appendix to this paper lays out the main categories of CSR issues employed by KLD in its rating process and also the average rating scores across industries.

¹² Sometimes a firm publishes multiple CSR reports, often discussing different CSR-related issues such as environmental versus non-environmental matters, in a single year. When that is the case, we combine them into one firm-year observation.

¹³ The statistics for page numbers are based on all CSR reports published in the year, not just on first-time reports.

TABLE 1
Sample Distribution

Panel A: Distribution by Industry

	Industries	No. of CSR Reports		Initial Sample		% of KLD Firms Disclosing CSR Reports	Average No. of Pages per Report	No. of First-Time CSR Reports	
		No. of CSR Reports	%	No. of CSR Reporters	%			No. of First-Time CSR Reports	%
1	Mining/Construction	33	2.77	9	3.06	10.84	43	7	3.29
2	Food	73	6.13	17	5.78	19.77	43	14	6.57
3	Textiles/Print/Publish	75	6.30	20	6.80	11.90	32	16	7.51
4	Chemicals	102	8.57	18	6.12	19.15	25	12	5.63
5	Pharmaceuticals	80	6.72	14	4.76	4.71	38	12	5.63
6	Extractive	96	8.07	16	5.44	8.89	33	13	6.10
7	Manf: Rubber/glass/etc.	18	1.51	4	1.36	7.55	47	3	1.41
8	Manf: Metal	28	2.35	4	1.36	4.00	30	3	1.41
9	Manf: Machinery	30	2.52	9	3.06	7.50	34	8	3.76
10	Manf: Electrical Eqpt	22	1.85	6	2.04	4.51	21	4	1.88
11	Manf: Transport Eqpt	62	5.21	13	4.42	14.94	41	10	4.69
12	Manf: Instruments	32	2.69	9	3.06	3.88	35	8	3.76
13	Manf: Misc.	3	0.25	2	0.68	8.00	53	2	0.94
14	Computers	123	10.34	27	9.18	4.55	36	23	10.80
15	Transportation	30	2.52	12	4.08	4.36	41	8	3.76
16	Utilities	188	15.80	45	15.31	30.41	39	29	13.62
17	Retail: Wholesale	22	1.85	5	1.70	4.72	24	3	1.41
18	Retail: Misc.	29	2.44	12	4.08	5.41	38	10	4.69
19	Retail: Restaurant	17	1.43	4	1.36	8.33	52	3	1.41
20	Financial	92	7.73	34	11.56	4.37	33	18	8.45
21	Insurance/Real Estate	1	0.08	1	0.34	0.40	29	0	0.00
22	Services	14	1.18	8	2.72	2.15	22	5	2.35
23	Others	20	1.68	5	1.70	20.00	48	2	0.94
	Total	1,190	99.99	294	99.97			213	100.01

(continued on next page)

Panel B: Distribution by Year

Year	No. of CSR Reports		%	No. of First-Time CSR Reporters		Initial Sample		No. of Firms in KLD Database	% of KLD Firms Disclosing CSR Reports	Average No. of Pages per Report	No. of First-Time CSR Reports		%
	No. of CSR Reports	%		No. of First-Time CSR Reporters	%	No. of First-Time CSR Reports	%						
1993	8	0.67		6	2.04	432	1.85	22	3	1.41			
1994	16	1.34		11	3.74	423	3.78	24	7	3.29			
1995	23	1.93		10	3.40	440	5.23	25	5	2.35			
1996	31	2.61		12	4.08	456	6.80	23	9	4.23			
1997	46	3.87		14	4.76	459	10.02	23	11	5.16			
1998	41	3.45		8	2.72	494	8.30	23	4	1.88			
1999	56	4.71		15	5.10	518	10.81	25	7	3.29			
2000	57	4.79		18	6.12	537	10.61	23	11	5.16			
2001	101	8.49		42	14.29	928	10.88	28	34	15.96			
2002	99	8.32		28	9.52	981	10.09	31	18	8.45			
2003	101	8.49		22	7.48	2,639	3.83	33	14	6.57			
2004	121	10.17		20	6.80	2,750	4.40	40	16	7.51			
2005	149	12.52		32	10.88	2,792	5.34	42	30	14.08			
2006	157	13.19		31	10.54	2,790	5.63	41	23	10.80			
2007	184	15.46		25	8.50	2,827	6.51	45	21	9.86			
Total	1,190	100.01		294	99.97				213	100.00			

This table provides the sample distribution by industry and year for both the initial CSR reports collected and CSR reports selected for the final sample. "No. of First-Time CSR Reports" is the number of CSR reports in the earliest reporting year of a firm (i.e., CSR report-initiating year) based on our collected CSR reports. "% of KLD Firms Disclosing CSR Reports" is the number of CSR disclosing firms in any industry (namely, "No. of CSR Reporters" in Panel A) or year (namely, "No. of CSR Reports" in Panel B) divided by the total number of KLD firms in that industry or year, respectively.

Empirical Models and Variable Definitions

Past Cost of Equity Capital and Current-Year CSR Disclosure

To test H1, we examine whether a high cost of equity capital in the previous year gives firms an incentive for CSR disclosure in the current year. In the empirical regression model, we control for other determinants of CSR disclosure to parse out potential confounding effects. However, the current literature provides limited information on what motivates a firm's CSR disclosure decision. As CSR disclosure is part of a firm's overall voluntary disclosure strategy, we identify potential factors from the voluntary disclosure literature that influence a firm's decision to commit to CSR disclosure. Our logistic regression model is specified as follows:

$$\begin{aligned} \log[\text{prob}(DISCI_{i,t})/(1 - \text{prob}(DISCI_{i,t}))] = & \beta_0 + \beta_1COC_{i,t-1} + \beta_2PERFORM_{i,t-1} \\ & + \beta_3HICONCERN_{i,t-1} + \beta_4SIZE_{i,t-1} \\ & + \beta_5LITIGATION_{i,t-1} + \beta_6ROA_{i,t-1} \\ & + \beta_7COMPETITION_{i,t-1} + \beta_8FIN_{i,t-1} \\ & + \beta_9TOBINQ_{i,t-1} + \beta_{10}LEV_{i,t-1} + \beta_{11}GLOBAL_{i,t-1} \\ & + \beta_{12}LIQUIDITY_{i,t-1} + \beta_{13}ABS\_EM_{i,t-1} \\ & + \beta_{14}CIG_{i,t-1} + \Sigma IND_{i,t} + \Sigma YEAR_{i,t} + \varepsilon_{i,t} \end{aligned} \quad (1)$$

where $DISCI_{i,t}$ is an indicator variable that equals 1 if firm i discloses a standalone CSR report for the first time in year t (initiating firm-years or initiators), and 0 (non-initiating firm-years or non-initiators) otherwise. Therefore, the control group ($DISCI = 0$), namely, non-initiators, includes all years of firms that never issue CSR reports and the years before and after CSR-initiating firms' first-time reports.

Our main variable of interest, the cost of equity capital in the year prior to first-time CSR disclosure, COC , is the *ex ante* or implied cost of equity capital, calculated using three different models, namely, those of Gebhardt et al. (2001), Claus and Thomas (2001), and Easton (2004). The mean of the three measures (COC_AVG) serves as our proxy for the cost of equity capital. To implement the estimation, we obtain expected future earnings per share from I/B/E/S and market price and dividend per share from Compustat.

We include a number of control variables in the regression. $PERFORM$ is the total KLD score of CSR strengths, which we use to proxy for firms' CSR performance. Firms with better social performance have a greater incentive to disclose (Dye 1985). The KLD database is widely used in CSR research (Graves and Waddock 1994; Berman et al. 1999; Baron et al. 2009). Waddock (2003, 369) regards it as "the *de facto* (CSR) research standard at the moment."¹⁴ KLD ranks firms' CSR performance in seven main categories: (1) community, (2) corporate governance, (3) diversity, (4) employee relations, (5) environment, (6) human rights, and (7) product.¹⁵ We adjust raw CSR strength scores each year by industry medians to get relative performance scores that are comparable across industries.

¹⁴ Of course, there is also no lack of criticism of the KLD database. For example, KLD uses indicator variables to describe firms' CSR performance. This is a crude methodology and potentially suffers from loss of information. Chatterji et al. (2009) show that KLD environmental strengths do not accurately predict pollution levels or compliance violations, and that KLD ratings do not optimally use publicly available data.

¹⁵ The rankings are based on information obtained from surveys, financial statements, government documents, peer-reviewed legal journals, and reports from mainstream media. KLD defines a set of potential strengths under each category and assigns a value of 1 if a strength exists, and a value of 0 otherwise. See the Appendix for more details on KLD's rating categories.

We control for firm size (*SIZE*) because size captures various factors motivating firms to issue CSR reports such as public pressure or financial resources (Lang and Lundholm 1993). We measure *SIZE* as the natural logarithm of the market value of common equity at the beginning of each year. Skinner (1997) argues that firms facing a higher level of litigation risk (*LITIGATION*) are more likely to make voluntary disclosure to preempt potential lawsuits. *LITIGATION* is an indicator variable that equals 1 if a firm operates in a high-litigation industry (SIC codes of 2833–2836, 3570–3577, 3600–3674, 5200–5961, and 7370), and 0 otherwise (Francis et al. 1994; Matsumoto 2002). As firms with better financial performance likely have more resources to practice CSR activities and produce CSR reports, we include return on assets (*ROA*), computed as income before extraordinary items scaled by total assets at the beginning of each year.

Dye (1985) suggests that proprietary costs arising from product market competition can reduce disclosure incentives. Hence, we control for industry competition (*COMPETITION*), which is proxied by the Herfindahl-Hirschman Index multiplied by -1 . This index is computed as the sum of the squared fractions of sales of the 50 largest firms in an industry (industries are defined based on the two-digit SIC codes). In cases where there are fewer than 50 firms in an industry, we use all firms in the industry to calculate market shares. In addition, firms raising capital in the public market have a greater propensity to make voluntary disclosures (Frankel et al. 1995). We control for a firm's financing activities (*FIN*) by assessing the amount of debt or equity capital raised by the firm during the year scaled by total assets at the beginning of the year. Following Richardson et al. (2004), *FIN* is measured as the sale of common and preferred shares minus the purchase of common and preferred shares plus the long-term debt issuance minus the long-term debt reduction.

We also control for growth opportunities (*TOBINQ*) because firms in an expansionary period are more financially constrained and have fewer resources for CSR activities and disclosure. However, growth firms also tend to have higher levels of information asymmetry, which could induce managers to make more disclosures to attract potential investors. The net effect is hence unknown *ex ante*. *TOBINQ* is Tobin's Q, defined as the market value of common equity plus the book value of preferred stock, book value of long-term debt and current liabilities, scaled by the book value of total assets. We include the debt ratio (*LEV*) in the model because debt servicing plays a monitoring role and debt holders demand greater disclosure (Leftwich et al. 1981). We define *LEV* as the ratio of total debt divided by total assets.

In addition, firms with a global focus, especially those operating in emerging markets, face greater pressure to commit to social performance and are accordingly more likely to provide CSR disclosure. *GLOBAL* is an indicator variable that equals 1 if a firm reports foreign income, and 0 otherwise. Further, managers have incentives to increase the liquidity of their firms' stock in order to issue equities or sell shares of their firm obtained from options or other incentive compensation plans. One way to increase liquidity is to improve transparency and supply more information to investors. Our liquidity measure, *LIQUIDITY*, is the ratio of the number of shares traded in the year to the total shares outstanding at the year-end.

Finally, CSR disclosure could be correlated with the general disclosure policies and financial transparency of firms. To control for this possibility, we include two variables to proxy for firm financial disclosure quality and voluntary disclosure policy: earnings quality (*ABS_EM*) and management earnings forecasts (*CIG*). We use the absolute value of abnormal accruals from the modified Jones (1991) model, based on Dechow et al. (1995), to proxy for earnings quality (Francis et al. 2008).¹⁶ Following prior studies that use management forecasts as a direct measure

¹⁶ Using the original Jones (1991) model or an alternative version developed by Dechow et al. (2003, 359, Equation (2b)) yields similar results.

of a firm's disclosure policy (Rogers and Van Buskirk 2009), we define *CIG* as an indicator variable that equals 1 if a firm issues at least one earnings forecast in the year, and 0 otherwise. In all specifications of the model, we include industry and year indicators to control for potential industry and year effects.

Effect of CSR Disclosure on the Future Cost of Equity Capital

Hypothesis 2 predicts that CSR disclosure leads to a lower cost of equity capital. We test H2 by estimating the following regression model:

$$\Delta\%COC_{i,t+1} = \beta_0 + \beta_1 DISCI_{i,t} + \beta_2 \Delta SIZE_{i,t} + \beta_3 \Delta BETA_{i,t} + \beta_4 \Delta LEV_{i,t} + \beta_5 \Delta MB_{i,t} + \beta_6 \Delta LTG_{i,t} + \beta_7 \Delta LNDISP_{i,t} + \Sigma IND_{i,t} + \Sigma YEAR_{i,t} + \varepsilon_{i,t} \quad (2)$$

where $\Delta\%COC_{i,t+1}$ is the percentage change in the cost of equity capital from year t to year $t+1$. The control variables also adopt the change form. A negative coefficient on *DISCI* would support H2.

The control variables are derived from prior research. Fama and French (1992) find that expected returns are negatively associated with firm size and positively associated with the book-to-market ratio. Hence, we include firm size (*SIZE*) and the market-to-book ratio (*MB*). The market model *BETA*, which is estimated using CRSP daily data for each year, is included to control for systematic risk. Gebhardt et al. (2001) and Gode and Mohanram (2003) find that the implied cost of equity capital is positively associated with long-term growth rate. We therefore include an empirical proxy of long-term growth rate based on I/B/E/S analyst EPS forecasts (*LTG*), which is measured as the difference between the two-year-ahead consensus EPS forecast and the one-year-ahead consensus EPS forecast scaled by the one-year-ahead consensus EPS forecast. Gebhardt et al. (2001) and Dhaliwal et al. (2005) find that analyst forecast dispersion is negatively associated with the implied cost of equity capital. Thus, we include analyst forecast dispersion (*LNDISP*), which is calculated as the logarithm of the standard deviation of analyst EPS forecasts divided by the consensus forecast. We include leverage (*LEV*) because Fama and French (1992) suggest that the cost of equity capital increases as the degree of leverage increases. All other variables are as defined earlier.

Although firms may be motivated by a possible reduction in the cost of equity capital when deciding whether to issue a CSR report, from the perspective of investors, CSR disclosure *per se* may not necessarily warrant a lower cost of equity capital. Corporate managers could attempt to manage public impressions through such disclosures; therefore, CSR information can be self-serving and noncredible (Cormier and Magnan 2003; Hobson and Kachelmeier 2005). Investors are likely to have a favorable perception if a firm actually performs well in its CSR practices relative to its peers. To incorporate this possibility, we augment Equation (2) with a measure of a firm's relative CSR performance from KLD (*HIPERFORM*):

$$\Delta\%COC_{i,t+1} = \beta_0 + \beta_1 DISCI_{i,t} + \beta_2 HIPERFORM_{i,t} + \beta_3 DISCI_{i,t} * HIPERFORM_{i,t} + \beta_4 \Delta SIZE_{i,t} + \beta_5 \Delta BETA_{i,t} + \beta_6 \Delta LEV_{i,t} + \beta_7 \Delta MB_{i,t} + \beta_8 \Delta LTG_{i,t} + \beta_9 \Delta LNDISP_{i,t} + \varepsilon_{i,t} \quad (3)$$

where *HIPERFORM* is an indicator variable that equals 1 if a firm's CSR performance score, *PERFORM*, is higher than its industry median (in other words, if the firm is a superior CSR performer in its industry), and 0 otherwise. All other variables are as defined earlier. We expect the effect of *DISCI * HIPERFORM* to be negative. In an additional test, instead of using the interaction term between *DISCI* and *HIPERFORM*, we estimate Equation (2) within the high and low partitions of CSR performance scores. Using partitioned subsamples sacrifices some power due to

reduced sample size, but has the benefit of flexibility that allows the effects of other variables to also vary based on high or low levels of CSR performance scores.

Endogeneity and self-selection could potentially affect our results. In our main analysis, we use a lead-lag approach to tackle these issues. To further enhance inferences based on our lead-lag approach, we adopt the Heckman and Hausman two-stage procedures and repeat our main analyses. The Heckman two-stage procedure introduces the inverse Mills ratio into the second-stage OLS regression to control for self-selection bias that is related to CSR disclosure. We obtain qualitatively similar results using the Heckman two-stage procedure.¹⁷ The Hausman test deals with potential endogeneity in the data. We conduct the Hausman test and find that endogeneity does not qualitatively affect our main results.

IV. RESULTS

Descriptive Statistics

Table 2, Panel A provides descriptive statistics for the variables included in Equation (1) for the full sample and separately for initiators and non-initiators. The cost of equity capital before CSR disclosure is significantly higher ($p = 0.04$) among CSR initiators (12.86 percent) than among non-initiators (11.98 percent). This difference is also reflected in a significantly positive correlation coefficient between *DISCI* and *COC_AVG* in Table 2, Panel B, providing initial support for H1.

Consistent with the theory on voluntary disclosure, firms voluntarily publishing standalone CSR reports tend to have superior CSR performance (*PERFORM*) relative to their industry peers. The difference in CSR performance between the two groups (1.613 for initiators versus -0.166 for non-initiators) is significant ($p < 0.01$). The correlation between *DISCI* and *PERFORM* is also significantly positive though at a relatively moderate level of 0.09 based on the Spearman correlation and 0.13 based on the Pearson correlation (see Table 2, Panel B). This highlights the importance of including *PERFORM* in our regression equations.

Initiators are significantly larger (*SIZE*: 9.147 for initiators versus 5.783 for non-initiators, $p < 0.01$) and more profitable (*ROA*: 0.051 for initiators versus 0.015 for non-initiators, $p < 0.01$) than non-initiators, lending support to the financial resources argument for CSR disclosure. Contrary to the proprietary information argument, initiators tend to observe greater industry competition than non-initiators (*COMPETITION*: -0.060 for initiators versus -0.069 for non-initiators, $p = 0.02$).

Initiators have a significantly lower level of financing than non-initiators (*FIN*: -0.019 for initiators versus 0.043 for non-initiators, $p < 0.01$). The negative financing level for initiators implies that these firms, in net effect, either have repurchased stock or redeemed their debts. Firms normally conduct repurchases when they believe that their stock is undervalued, indicating a high cost of equity capital, which in turn provides an incentive for managers to increase disclosure and transparency levels. Similarly, the redemption of mature debts likely implies that firms need future financing to maintain a normal capital level. These firms would also be willing to increase their level of disclosure if doing so helped them to lower the cost of borrowing.

Initiators have a higher degree of leverage than non-initiators (*LEV*: 0.265 for initiators versus 0.221 for non-initiators, $p < 0.01$). Those with a higher level of global operations are also more likely to publish CSR reports (*GLOBAL*: 0.460 for initiators versus 0.219 for non-initiators, $p < 0.01$), consistent with the notion that these firms attract more attention in the international community. Contrary to the notion of disclosing information to improve liquidity, initiators actually

¹⁷ The only exception is for the test of analyst forecast errors. Among better CSR performers, the coefficient on *DISCI* is positive and insignificant.

TABLE 2
Summary Statistics and Correlation

Panel A: Mean Comparison		Full Sample			$DISCI_t = 1$ (n = 213)	$DISCI_t = 0$ (n = 11,712)	t-value (difference)
Variable		11.988			12.855	11.985	2.08
$COC\_AVG_{t-1}(\%)$		-0.142			1.613	-0.166	11.66
$PERFORM_{t-1}$		5.795			9.147	5.783	41.13
$SIZE_{t-1}$		0.204			0.199	0.204	-0.22
$LITIGATION_{t-1}$		0.017			0.051	0.015	13.06
ROA_{t-1}		-0.069			-0.060	-0.069	2.28
$COMPETITION_{t-1}$		0.042			-0.019	0.043	-14.36
FIN_{t-1}		2.046			1.994	2.047	-0.53
$TOBINQ_{t-1}$		0.221			0.265	0.221	4.70
LEV_{t-1}		0.192			0.460	0.219	5.44
$GLOBAL_{t-1}$		1.248			1.387	1.247	2.10
$LIQUIDITY_{t-1}$		0.528			0.646	0.527	15.09
ClG_{t-1}		0.068			0.032	0.066	-6.84
ABS_EM_{t-1}							

Panel B: Spearman/Pearson correlation (n ($DISCI = 0$) = 11,712, n ($DISCI = 1$) = 213)		$DISCI_t$	COC_AVG_{t-1}	$PERFORM_{t-1}$	$SIZE_{t-1}$	$LITIGATION_{t-1}$	ROA_{t-1}	$COMPETITION_{t-1}$
$DISCI_t$			0.01	0.13	0.11	0.00	0.02	0.01
COC_AVG_{t-1}	0.01			0.00	0.01	-0.07	-0.11	0.03
$PERFORM_{t-1}$	0.09		0.01		0.23	0.00	0.07	-0.03
$SIZE_{t-1}$	0.08		0.01	0.21		-0.02	0.25	0.09
$LITIGATION_{t-1}$	0.00		-0.09	-0.03	-0.03		-0.16	0.10
ROA_{t-1}	0.02		-0.06	0.04	0.24	-0.03		-0.09
$COMPETITION_{t-1}$	0.01		0.07	-0.03	0.12	0.07	-0.18	
FIN_{t-1}	-0.02		-0.11	-0.06	-0.14	0.08	-0.09	0.02

(continued on next page)

Panel B: Spearman/Pearson correlation ($n(DISCI = 0) = 11,712$, $n(DISCI = 1) = 213$)

	$DISCI_t$	COC_AVG_{t-1}	$PERFORM_{t-1}$	$SIZE_{t-1}$	$LITIGATION_{t-1}$	ROA_{t-1}	$COMPETITION_{t-1}$
$TOBINQ_{t-1}$	0.00	-0.32	0.02	0.23	0.18	0.29	0.10
LEV_{t-1}	0.02	0.15	0.01	0.08	-0.15	-0.15	-0.03
$GLOBAL_{t-1}$	0.05	-0.05	0.03	0.24	0.03	0.10	-0.07
$LIQUIDITY_{t-1}$	0.02	-0.06	-0.01	0.28	0.22	0.05	-0.04
CIG_{t-1}	0.08	0.02	0.08	0.23	0.04	0.10	0.00
ABS_EM_{t-1}	-0.03	-0.07	-0.02	-0.21	0.05	-0.04	-0.06

	FIN_{t-1}	$TOBINQ_{t-1}$	LEV_{t-1}	$GLOBAL_{t-1}$	$LIQUIDITY_{t-1}$	CIG_{t-1}	ABS_EM_{t-1}
$DISCI_t$	-0.02	0.00	0.01	0.03	0.01	0.08	-0.03
COC_AVG_{t-1}	-0.09	-0.22	0.14	0.01	-0.01	0.01	-0.04
$PERFORM_{t-1}$	-0.06	0.02	0.01	0.03	-0.02	0.09	-0.04
$SIZE_{t-1}$	-0.14	0.15	0.04	0.12	0.22	0.22	-0.20
$LITIGATION_{t-1}$	0.11	0.17	-0.13	0.01	0.21	0.04	0.05
ROA_{t-1}	-0.28	-0.15	-0.06	0.04	-0.05	0.08	-0.20
$COMPETITION_{t-1}$	0.03	0.08	-0.04	0.02	0.04	0.02	-0.02
FIN_{t-1}		0.30	-0.27	-0.04	0.08	-0.07	0.14
$TOBINQ_{t-1}$	0.18		-0.25	-0.03	0.29	0.03	0.21
LEV_{t-1}	-0.26	-0.37		-0.01	-0.06	0.00	-0.05
$GLOBAL_{t-1}$	-0.05	0.01	-0.01		0.04	0.07	-0.05
$LIQUIDITY_{t-1}$	0.08	0.32	-0.07	0.12		0.17	0.14
CIG_{t-1}	-0.06	0.08	0.01	0.14	0.22		-0.06
ABS_EM_{t-1}	0.10	0.14	-0.11	-0.07	0.11	-0.06	

All continuous variables are winsorized at the 1st and 99th percentiles. In Panel B, the Spearman (Pearson) correlations are below (above) the diagonal. A correlation coefficient in bold indicates that the correlation is statistically significant at least at the 10 percent level.

Variable Definitions:

$DISCI_t$ = indicator variable that equals 1 if it is the earliest reporting year of a firm that issues CSR reports (CSR report initiating year), and 0 otherwise;

COC_AVG_{t-1} = implied cost of equity capital (in percentage) in year $t-1$ estimated as the mean of three different models: Gebhardt et al. (2001), Claus and Thomas (2001), and Easton (2004);

(continued on next page)

- $PERFORM_{t-1}$ = measure of social performance defined as the industry-adjusted total CSR strength scores from the seven CSR rating categories obtained from the KLD Research & Analytics database;
- $SIZE_{t-1}$ = natural logarithm of the market value of equity ($CSHO * PRCC_F$) at the beginning of each year;
- $LITIGATION_{t-1}$ = indicator variable that equals 1 if the firm operates in a high-litigation industry (SIC codes of 2833–2836, 3570–3577, 3600–3674, 5200–5961, and 7370), and 0 otherwise;
- ROA_{t-1} = total return on assets measured as the ratio of income before extraordinary items (IB) over total assets (AT) at the beginning of each year;
- $COMPETITION_{t-1}$ = Herfindahl-Hirschman Index multiplied by -1 . The Herfindahl-Hirschman Index is calculated by summing the squares of the market shares of the 50 largest companies in an industry. We calculate a firm's market share by dividing the sales (SALE) of a firm in year t by the total sales of all of the 50 largest companies in an industry in that year. We define industries based on the two-digit SIC codes. In cases where there are fewer than 50 companies in an industry, we use all companies in that industry to calculate the market share of each firm;
- FIN_{t-1} = amount of debt or equity capital raised by the firm scaled by total assets at the beginning of year $t-1$. It is measured as the issuance of common stock and preferred shares minus the purchase of common stock and preferred shares ($SSTK_t - PRSTKC$) plus the long-term debt issuance minus the long-term debt reduction ($DLTIS_t - DLTR_t$);
- $TOBINQ_{t-1}$ = market value of common equity plus the book value of preferred stock (PSTKL), book value of long-term debt (DLTT) and current liability (LCT), scaled by the book value of total assets;
- LEV_{t-1} = leverage ratio, which is defined as the ratio of total debt (DLTT + DLC) divided by total assets;
- $GLOBAL_{t-1}$ = indicator variable that equals 1 if the firm reports non-zero foreign income (PIFO), and 0 otherwise;
- $LIQUIDITY_{t-1}$ = ratio of the number of shares traded in year $t-1$ to the total shares outstanding at the end of year $t-1$;
- CIG_{t-1} = indicator variable that equals 1 if the firm issues an earnings forecast in year $t-1$, and 0 otherwise. We obtain data for CIG from First Call database; and
- ABC_EM_{t-1} = absolute value of abnormal accruals estimated based on the modified Jones model. The modified-Jones model discretionary accrual is an estimated cross-sectional each year using all firm-year observations in the same two-digit SIC code as follows: $TA_{i,t} = \beta_0 + \beta_1(1 / Assets_{i,t-1}) + \beta_2(\Delta Sales_{i,t} - \Delta REC_{i,t}) + \beta_3 PPE_{i,t} + \varepsilon_{i,t}$, where $TA_{i,t}$ is defined as the change in non-cash current assets minus the change in current liabilities excluding the current portion of long-term debt, minus depreciation and amortization, scaled by lagged total assets; $\Delta Sales_{i,t}$ is the change in sales scaled by lagged total assets; $\Delta REC_{i,t}$ is the change in accounts receivable scaled by lagged total assets; and $PPE_{i,t}$ is net property, plant, and equipment scaled by lagged total assets. Only industry-years with at least eight firms for estimation are considered. We use the residuals from the annual cross-sectional industry regression model above as discretionary accruals.

have higher liquidity levels than non-initiators (*LIQUIDITY*: 1.387 for initiators versus 1.247 for non-initiators, $p = 0.04$). Finally, initiators have better financial disclosure as manifested in their more frequent management forecasts (*CIG*: 0.646 for initiators versus 0.527 for non-initiators, $p < 0.01$) and better earnings quality (*ABS_EM*: 0.032 for initiators versus 0.066 for non-initiators, $p < 0.01$) than non-initiators.

Cost of Equity Capital and the Likelihood of CSR Disclosure

Hypothesis 1 predicts that a firm's likelihood of disclosing its corporate social responsibility activities is positively associated with its cost of equity capital in the previous year. We report the regression results for Equation (1) in Table 3. In Column I, we include all first-time reporting firm-year observations. In Column II, we exclude first-time reports that primarily discuss environmental issues, following Simnett et al. (2009). In Column III, we examine the robustness of our results to the exclusion of the Utilities industry.

Across all three specifications of the dependent variable, the cost of equity capital, *CO-C_AVG*, in year $t-1$, is significantly positively associated with a firm's likelihood of voluntarily issuing a standalone CSR report in year t (coeff. = 0.049, $p < 0.01$; coeff. = 0.052, $p < 0.01$; and coeff. = 0.062, $p < 0.01$ in Columns I, II and III, respectively), consistent with H1, which posits that a higher past cost of equity capital is associated with a greater likelihood of voluntary CSR disclosure in the current year. In Column I, for instance, holding other factors constant, when the prior year cost of equity capital increases by one percentage point, the odds of initiating standalone CSR disclosure increase by 5.02 percent.

The coefficient estimates of the control variables are generally consistent with the univariate comparisons in Table 2. One exception is *LIQUIDITY*, which reverses direction. The significantly negative coefficient suggests that firms with lower levels of liquidity are more likely to publish CSR reports, consistent with our original conjecture. The effects of financial disclosure quality, *ABS_EM*, and management forecast, *CIG*, are no longer significant.

CSR Disclosure and the Future Cost of Equity Capital

Hypothesis 2 predicts that voluntary CSR disclosure leads to a lower future cost of equity capital. Table 4, Panel A compares initiators and non-initiators and Table 4, Panel B presents the regression results. In Column I (Equation (2)), the coefficient on *DISCI* is insignificant (coeff. = -0.037 , $p > 0.50$). It appears that CSR disclosure *per se* is not significantly associated with a change in a firm's future cost of equity capital. In Column II (Equation (3)), we consider whether a firm has superior CSR performance relative to its industry peers. The interaction term between *DISCI* and *HIPERFORM* is significantly negative (coeff. = -4.618 , $p < 0.01$), consistent with H2, which posits that CSR disclosure reduces the cost of equity capital.¹⁸ Combining the main effect of *DISCI* and the effect of the interaction term between *DISCI* and *HIPERFORM* in Column II, we infer that superior CSR performers enjoy a 1.833 percent reduction in the cost of equity capital when they produce standalone CSR reports for the first time. In Columns III and IV, we obtain similar results when we exclude environmental reports and the Utilities industry, respec-

¹⁸ We perform a sensitivity test by restricting the analysis to firm-year observations of CSR reporters in a pre-post setting. Specifically, we focus only on disclosing firms and still use the change specification of the dependent variable $\Delta CO-C\_AVG$. *DISCI* is an indicator variable that equals 1 for the first reporting year and equals 0 before or after the first reporting year of a disclosing firm. The purpose of this examination is to show that a reduction in the cost of equity capital occurs immediately after the first reporting year and to alleviate the concern that the size of initiator sample is small relative to the universe of firm-year observations used in the main test. We obtain similar results and inferences, namely, firms with superior CSR performance enjoy a reduction in the cost of equity capital if they publish standalone CSR reports.

TABLE 3
Determinants of CSR Disclosure

Dependent Variable = $DISCI_t$	Full Sample		Removing Environmental Reports		Removing Utilities		
Variables	Pred. Sign	Coeff.	Prob.	Coeff.	Prob.	Coeff.	Prob.
COC_AVG_{t-1}	+	0.049***	0.00	0.052***	0.00	0.062***	0.00
$PERFORM_{t-1}$	+	0.101**	0.03	0.057	0.25	0.100**	0.04
$SIZE_{t-1}$	+	0.967***	0.00	1.017***	0.00	1.001***	0.00
$LITIGATION_{t-1}$	?	0.208	0.45	0.154	0.59	0.187	0.50
ROA_{t-1}	+	1.404	0.47	1.407	0.49	1.805	0.36
$COMPETITION_{t-1}$	-	1.685	0.37	1.105	0.55	1.569	0.40
FIN_{t-1}	+	0.768	0.50	0.955	0.41	1.406	0.20
$TOBINO_{t-1}$	?	-0.398***	0.00	-0.355***	0.01	-0.371***	0.00
LEV_{t-1}	+	-0.846	0.28	-0.851	0.30	-1.038	0.20
$GLOBAL_{t-1}$	+	0.078*	0.08	0.091**	0.05	0.084*	0.07
$LIQUIDITY_{t-1}$	-	-0.292**	0.02	-0.255**	0.04	-0.292**	0.02
CIG_{t-1}	+	0.313	0.16	0.324	0.18	0.272	0.27
ABS_EM_{t-1}	-	-2.217	0.49	-1.545	0.64	-3.137	0.36
Year Indicators		Yes		Yes		Yes	
Industry Indicators		Yes		Yes		Yes	
Pseudo R ²		0.258		0.244		0.271	
Likelihood Ratio		303.40		27.00		273.30	
n (dep. var. = 1)		213		164		182	
n		11,925		11,876		11,419	

*, **, *** Indicate that the estimated coefficient is statistically significant at the 10 percent, 5 percent, and 1 percent levels, respectively.

This table presents the logistic regression results with $DISCI_t$ in year t as the dependent variable, while all control variables are in year $t-1$. All continuous variables are winsorized at the 1st and the 99th percentiles. All t-statistics are corrected using the Huber-White procedure.

All variables are defined as in Table 2.

TABLE 4
Post-CSR Disclosure Cost of Equity Capital

Panel A: Mean Comparison

	Full Sample	$DISCI_t = 1$	$DISCI_t = 0$	t-value (difference)
$\Delta COC\_AVG_{t+1}(\%)$	1.691	1.393	1.693	-0.90
$\Delta SIZE_t$	-0.003	-0.002	-0.003	0.13
$\Delta BETA_t$	-0.019	0.061	-0.020	2.79
ΔLEV_t	-0.023	-0.027	-0.023	-4.01
ΔMB_t	-0.015	-0.012	-0.015	0.85
ΔLTG_t	0.004	-0.051	0.004	-1.10
$\Delta LNDISP_t$	0.023	0.009	0.023	-1.47

Panel B: Post-CSR Disclosure Cost of Equity Capital (Dependent Variable = ΔCOC_AVG_{t+1})

Variables	I		II		III		IV	
	Coeff.	t-stat.	Full Sample		Removing Environmental Reports		Removing Utilities	
$DISCI_t$	-0.037	-0.06	2.785***	2.54	4.699***	3.56	1.856	1.45
$HIPERFORM_t$			0.030	0.20	0.041	0.27	0.052	0.33
$DISCI_t * HIPERFORM_t$			-4.618***	-3.38	-6.015***	-3.74	-4.045***	-2.61
$\Delta SIZE_t$	4.698***	3.70	4.849***	3.82	4.814***	3.79	5.040***	3.86
$\Delta BETA_t$	0.342***	3.53	0.331***	3.40	0.334***	3.44	0.298***	2.99
ΔLEV_t	-7.220***	2.86	-7.107***	2.81	-5.116**	-2.04	6.557***	2.54
ΔMB_t	-10.137***	-14.19	-10.190***	-14.23	-10.257***	-14.33	-10.107***	-13.86
ΔLTG_t	-0.175**	-2.45	-0.173**	-2.42	-0.170**	-2.37	-0.175**	-2.37
$\Delta LNDISP_t$	0.111	0.30	0.107	0.29	0.109	0.30	0.243	0.64
Year Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R ²	0.062	0.062	0.064	0.064	0.062	0.062	0.062	0.062

(continued on next page)

Panel B: Post-CSR Disclosure Cost of Equity Capital (Dependent Variable = $\Delta\text{COC\_AVG}_{t+1}$)

Variables	I		II		III		IV	
	Coeff.	t-stat.	Full Sample		Removing Environmental Reports		Removing Utilities	
n ($\text{DISCI}_t = 1$)	213		Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.
n ($\text{DISCI}_t = 1$ and $\text{HIPERFORM}_t = 1$)			213		164		184	
n	9,254		165		127		150	
			9,254		9,205		8,820	

Panel C: Post-CSR Disclosure Cost of Equity Capital Partitioned Conditional on Firm's CSR Performance

Variables	Top 50% PERFORM.		Bottom 50% PERFORM.	
	Coeff.	t-stat.	Coeff.	t-stat.
DISCI_t	-1.777**	-2.04	2.130	1.41
ΔSIZE_t	3.200***	6.40	5.673***	3.99
ΔBETA_t	0.360***	3.12	0.239	0.79
ΔLEV_t	-4.857*	-1.68	-7.747	-1.11
ΔMB_t	-12.258***	-14.16	-4.778**	-2.13
ΔITG_t	-0.332***	-3.87	-0.478**	-2.17
ΔINDISP_t	0.414	0.93	-0.960	-0.78
Year Indicators		Yes		Yes
Industry Indicators		Yes		Yes
Adjusted R ²		0.048		0.059
n ($\text{DISCI} = 1$)		153		60
n		4,691		4,563

*, **, *** Indicate that the estimated coefficient is statistically significant at the 10 percent, 5 percent and 1 percent levels, respectively.

All continuous variables are winsorized at the 1st and 99th percentiles. All t-statistics are corrected using the Huber-White procedure. Analyst forecast data are obtained from I/B/E/S.

Variable Definitions:

DISCI_t = indicator variable that equals 1 if it is the earliest reporting year of a firm that issues CSR reports (CSR report initiating year), and 0 otherwise;

$\Delta\text{COC\_AVG}_{t+1}$ = change in cost of equity capital (in percentage) from year t to year $t+1$;

(continued on next page)

$HIPERFORM_t$ = indicator variable that equals 1 if the firm's total CSR strength score, namely, $PERFORM$ in year t , is higher than the industry median (in other words, if the firm is classified as a high CSR performer in its industry), and 0 otherwise;
 $\Delta SIZE_t$ = change in size where $SIZE$ is measured as the natural logarithm of the market value of equity ($CSHO * PRCC\_F$) at the beginning of each year;
 $\Delta BETA_t$ = change of beta from year $t-1$ to year t where $BETA$ is estimated from the market model using the daily CRSP stock returns;
 ΔLEV_t = change in leverage from year $t-1$ to year t where LEV is the leverage ratio, which is defined as the ratio of total debt ($DLTT + DLC$) divided by total assets;
 ΔMB_t = change of market-to-book ratio from year $t-1$ to year t where the market-to-book ratio is defined as the market value of equity divided by book value of equity (CEQ);
 ΔLTG_t = long-term growth rate where LTG is measured as the difference between the mean two-year-ahead analyst consensus EPS forecast and the mean one-year-ahead analyst consensus EPS forecast divided by the mean one-year-ahead analyst consensus EPS forecast; and
 $\Delta LNDISP_t$ = change in analyst forecast dispersion from year $t-1$ to year t , where $LNDISP$ is measured as the logarithm of the standard deviation of analyst estimates of year t earnings divided by the consensus forecast of year t earnings.

tively. Overall, the evidence is consistent with our H2 that CSR-disclosing firms with superior CSR performance achieve a reduction in the cost of equity capital.¹⁹

Table 4, Panel C presents the results from estimating Equation (2) within the two subsamples partitioned based on annual industry medians of CSR performance (*PERFORM*). Consistent with the results in Panel B, we find a significantly negative coefficient on *DISCI* (coeff. = -1.777 , $p = 0.04$) in the high CSR performance subsample. This coefficient indicates that voluntary CSR disclosure yields a 1.77 percent reduction in the cost of equity capital. In the low CSR performance subsample, there is no significant association between CSR disclosure and the change in the cost of equity capital.²⁰

Potential Mechanisms Linking CSR Disclosure and the Cost of Equity Capital

The above results suggest that CSR disclosure combined with superior CSR performance is associated with a reduction in the cost of equity capital. Below, we provide evidence on the potential underlying mechanisms through which voluntary CSR disclosure lowers the cost of equity capital. We focus on two types of financial intermediaries: institutional investors and financial analysts.

CSR Disclosure and Institutional Investors

Shleifer and Vishny (1986) suggest that the large equity stakes in the invested firms and the high levels of sophistication of these investors enable them to reduce agency cost problems and the extent of information asymmetry between managers and shareholders, an effect that leads to a reduction in the cost of equity capital. We consider three different types of institutional investors: dedicated (*DED*), transient (*TRA*), and quasi-indexer (*QIX*) institutional investors. Dedicated institutional investors are more likely to play monitoring and governance roles than the other two types (Bushee 1998). To determine whether CSR disclosure attracts institutional investors, we follow Bushee and Noe (2000) and estimate the following model:

$$\begin{aligned} \Delta INST_{i,t+1} = & \beta_0 + \beta_1 DISCI_{i,t} + \beta_2 HIPERFORM_{i,t} + \beta_3 DISCI_{i,t} * HIPERFORM_{i,t} + \beta_4 INST_{i,t-1} \\ & + \beta_5 \Delta MRET_{i,t} + \beta_6 TVOL_{i,t-1} + \beta_7 \Delta MV_{i,t} + \beta_8 BETA_{i,t-1} + \beta_9 IRISK_{i,t-1} \\ & + \beta_{10} \Delta LEV_{i,t} + \beta_{11} \Delta DP_{i,t} + \beta_{12} \Delta EP_{i,t} + \beta_{13} \Delta MB_{i,t} + \beta_{14} \Delta SGR_{i,t} + \beta_{15} \Delta RATE_{i,t} \\ & + \beta_{16} \Delta SHRS_{i,t} + \Sigma IND_{i,t} + \Sigma YEAR_{i,t} + \varepsilon_{i,t}, \end{aligned} \quad (4)$$

¹⁹ In alternative specifications of the model, we examine the effects of two other variables proxying for firms' effort and commitment to better CSR disclosure. (1) We identify firms that provide assurance (*ASSURANCE*) of their reports through independent third parties, most often Big 4 accounting firms and international consulting companies. Simnett et al. (2009) provide evidence that firms seeking to enhance the creditability of their reports and their corporate reputation are more likely to have their sustainability reports assured. (2) We also assess the effect of the length of each CSR report (*LENGTH*) relative to the average report length of the disclosing firm's industry (Leuz and Schrand 2008). Of course, *ASSURANCE* and *LENGTH* are not independent of *DISCI*. We find that, conditional on first-time CSR disclosure (*DISCI*), external assurance and long report length further reduce the cost of equity capital. Specifically, when we use the *ASSURANCE* indicator (equals 1 with an assurance, and 0 otherwise), the coefficient on *DISCI * HIPERFORM* is negative and significant (coeff. = -3.523 , $p < 0.01$) and the coefficient on *DISCI * HIPERFORM * ASSURANCE* is negative and significant (coeff. = -3.540 , $p = 0.08$). Therefore, assurance doubles the effect of CSR disclosure. When we use the *LENGTH* indicator (equals 1 if longer than the industry-year median, and 0 otherwise), the coefficient on *DISCI * HIPERFORM* is negative and significant (coeff. = -2.574 , $p = 0.02$) and the coefficient on *DISCI * HIPERFORM * LENGTH* is negative and significant (coeff. = -3.930 , $p = 0.06$). Therefore, a long report more than doubles the effect of CSR disclosure.

²⁰ The coefficient on *DISCI* is not significant but positive, if anything, for poor CSR performers. It is possible that disclosing poor CSR performance could actually be a signal of high risk or firm weakness and, therefore, the cost of equity capital could actually go up. This does explain why the direct effect of *DISCI* is insignificant in the pooled regression.

where Δ denotes a change from year t to year $t+1$. *INST* represents stock ownership by dedicated (*DED*), transient (*TRA*), or quasi-indexer (*QIX*) institutional investors. *MRET* is the market-adjusted buy-and-hold stock return measured over the year. *TVOL*, a liquidity proxy, is the average monthly trading volume relative to total shares outstanding. *IRISK* is the logarithmic transformation of the standard deviation of market-model residuals calculated using daily stock returns. Beta (*BETA*), debt ratio (*LEV*), and *IRISK* capture firm risk along different dimensions. *DP* is the ratio of dividends to the market value of equity. *EP* is the ratio of income before extraordinary items to the market value of equity. *SGR* is the percentage change in annual sales. We include *DP*, *EP*, *MB*, and *SGR* to control for changes in firms' fundamentals that can affect the investment decisions of institutional investors (Bushee 2001). *RATE* is the S&P stock rating (9 = A+, 8 = A, 7 = A-, 6 = B+, 5 = B, 4 = B-, 3 = C, 2 = D, 1 = not rated), which captures the preference of institutional investors for well-reputed firms (Del Guercio 1996). *SHRS* is the logarithmic transformation of shares outstanding, and its change form proxies for equity issuance or repurchases that affect both institutional investor following and firms' disclosure policies. All other variables are as defined earlier.

Table 5, Panel A presents comparisons of one-year-ahead holdings and changes in holdings by the three types of institutional investors between initiators and non-initiators. Overall, the univariate comparisons do not reveal significant differences between initiators and non-initiators. If anything, we observe a greater decrease in transient institutional holding among initiators compared to non-initiators ($p = 0.04$), even though the level of this type of holding is still slightly higher among initiators than among non-initiators ($p = 0.07$).

Table 5, Panel B displays the regression results. There is weak evidence that initiating firms with superior CSR performance attract more dedicated institutional investors. The coefficient on *DISCI * HIPERFORM* is marginally significantly positive (coeff. = 0.414, $p = 0.16$). To further examine this issue, we run regressions without the interaction term in the two subsamples partitioned based on annual industry medians of CSR performance for dedicated institutional investors. We report the results in Table 5, Panel C. We observe a significantly positive coefficient on *DISCI* (coeff. = 0.438, $p = 0.01$) for the superior-performance group, whereas the coefficient on *DISCI* for the low-performance group is insignificant. In untabulated tests, we do not find a significant association between transient or quasi-indexer institutional investor holdings and the initiation of CSR disclosure for the full sample or the partitioned subsamples.

In sum, the evidence in this subsection suggests that voluntary CSR disclosure attracts dedicated institutional investors, who have long investment horizons and play monitoring and governance roles. Consistent with our previous evidence that superior CSR performers enjoy a reduction in the cost of equity capital through CSR disclosure, the effect of CSR disclosure on dedicated institutional ownership is stronger if disclosing firms have CSR performance superior to their industry peers.

CSR Disclosure and Analyst Forecasts

We also examine three questions related to financial analysts and CSR disclosure. First, we explore whether financial analysts are more willing to cover firms after they initiate CSR disclosure. Second, we investigate whether the level of forecast accuracy increases and finally we determine whether forecast dispersion decreases when CSR reports are available. Increased levels of analyst coverage and forecast accuracy and a reduction in the level of forecast dispersion have the potential to lower the cost of equity capital. To determine the impact of CSR disclosure on the behavior of financial analysts, we run the following three regressions following Lang and Lundholm (1996) and Ali et al. (2007):

TABLE 5
Post-CSR Disclosure Changes in Institutional Ownership

Panel A: Mean Comparison

	Full Sample	$DISC_{it} = 1$	$DISC_{it} = 0$	t-value (difference)
$DED_{t+1}(\%)$	11.996	11.285	12.019	-0.79
$TRA_{t+1}(\%)$	16.507	18.572	16.440	1.82
$QIX_{t+1}(\%)$	29.809	30.014	29.802	0.24
ΔEDD_{t+1}	-0.551	-0.995	-0.536	-0.57
ΔTRA_{t+1}	-0.452	-2.293	-0.390	-2.07
ΔQIX_{t+1}	1.018	1.947	0.986	1.00

Panel B: Post-CSR Disclosure Institutional Investor Holdings

Variables	ΔEDD_{t+1}		ΔQIX_{t+1}		ΔTRA_{t+1}	
	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.
$DISC_{it}$	0.190	0.80	0.267	0.63	0.053	0.16
$HIPERFORM_{it}$	-0.037	-0.76	-0.129	-1.44	0.085	1.23
$DISC_{it} * HIPERFORM_{it}$	0.414	1.40	-0.093	-0.17	-0.052	-0.13
DED_{t-1}	-0.018***	-5.94	-0.032***	-6.75	-0.059***	-15.37
QIX_{t-1}					-0.010	-0.13
TRA_{t-1}					0.066	1.24
$\Delta MRET_{it}$	0.043	0.82	-0.107	-1.07	-0.149	-1.08
$TVOL_{t-1}$	0.019	0.57	0.033	0.49	-0.082	-0.85
ΔMV_{it}	0.071	0.74	0.026	0.14	-0.087	-0.74
$BETA_{t-1}$	-0.053	-0.78	-0.010	-0.08	0.324	0.63
$IRISK_{t-1}$	0.056	0.69	0.065	0.42	6.617**	1.69
ΔLEV_{it}	-0.008	-0.02	-0.924	-1.40	-0.356	-0.88
ΔDP_{it}	4.032	1.52	11.935**	2.39		
ΔEP_{it}	0.169	0.61	0.152	0.29		

(continued on next page)

Panel B: Post-CSR Disclosure Institutional Investor Holdings

Variables	ΔEDD_{t+1}		ΔQIX_{t+1}		ΔTRA_{t+1}	
	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.
ΔMB_t	-0.010	-1.14	0.011	0.66	0.026**	1.96
ΔSGR_t	-0.001	-0.22	0.002	0.87	-0.001	-0.18
$\Delta RATE_t$	0.034***	2.79	0.037*	1.66	0.023	1.35
$\Delta SHRS_t$	0.298**	2.34	-0.109	-0.47	-0.287	-1.60
Year Indicators				Yes		Yes
Industry Indicators				Yes		Yes
Adjusted R ²		0.747		0.792		0.724
n ($DISCI_t = 1$)		170		170		170
n ($DISCI_t = 1$ and $HIPERFORM_t = 1$)		126		126		126
n		9,342		9,342		9,342

Panel C: Post-CSR Disclosure Institutional Investor Holdings Conditional on Firm's CSR Performance

Variables	TOP 50% PERFORM		Bottom 50% PERFORM	
	Coeff.	t-stat.	Coeff.	t-stat.
$DISCI_t$	0.438***	2.67	-0.063	-0.24
DED_{t-1}	-0.017***	-3.91	-0.027***	-6.45
$\Delta MRET_t$	0.047	0.60	-0.031	-0.42
$TVOL_{t-1}$	0.028	0.53	0.066	1.41
ΔMV_t	0.145	1.01	0.050	0.38
$BETA_{t-1}$	-0.151	-1.57	0.039	0.39
$IRISK_{t-1}$	0.178	1.57	-0.126	-1.06
ΔLEV_t	0.472	0.90	-0.199	-0.42
ΔDP_t	5.020	1.27	3.690	1.01
ΔEP_t	-0.104	-0.26	0.427	1.08
ΔMB_t	-0.012	-0.97	0.009	0.68

(continued on next page)

Panel C: Post-CSR Disclosure Institutional Investor Holdings Conditional on Firm's CSR Performance

Variables	TOP 50% PERFORM		Bottom 50% PERFORM	
	Coeff.	t-stat.	Coeff.	t-stat.
ΔSGR_t	-0.004**	-2.42		
$\Delta RATE_t$	0.042***	2.49		
$\Delta SHRS_t$	0.457***	2.55		
Year Indicators				
Industry Indicators				
Adjusted R ²				
n ($DISCI_t = 1$)				
n				

*, **, *** Indicate that the estimated coefficient is statistically significant at the 10 percent, 5 percent, and 1 percent levels, respectively. All continuous variables are winsorized at the 1st and 99th percentiles. All t-statistics are corrected using the Huber-White procedure. Panel A presents the OLS regression results with ΔDEB_{t+1} , ΔQIX_{t+1} , or ΔTRA_{t+1} as the dependent variable.

Variable Definitions:

$DISCI_t$ = indicator variable that equals 1 if it is the earliest reporting year of a firm that issues CSR reports (CSR report initiating year), and 0 otherwise;
 DEB_{t+1} , QIX_{t+1} , and TRA_{t+1} = percentage of ownership holding by dedicated, quasi-indexer, and transient institutional investors, respectively, relative to total shares outstanding in year $t+1$;
 ΔDEB_{t+1} , ΔQIX_{t+1} , and ΔTRA_{t+1} = changes in percentage of ownership from year t to year $t+1$. If there is a missing value on the percentage holding of any type of institutional investor when a firm's total institutional investor holding does not equal 0, then we assume a value of 0;
 $HIPERFORM_t$ = indicator variable that equals 1 if the firm's total CSR strength score, namely, $PERFORM$, is higher than the industry median (in other words, if the firm is classified as a high CSR performer in its industry), and 0 otherwise;
 $\Delta MRET_t = MRET_t - MRET_{t-1}$; where $MRET$ is the market-adjusted buy-and-hold stock return measured over the year of interest with a minimum of 125 observations;
 $TVOL_{t+1}$ = average monthly trading volume relative to total shares outstanding measured over the year of interest;
 $\Delta MV_t = MV_t - MV_{t-1}$; where MV is the logarithm of the market value of equity ($CSHO * PRCC_F$);
 $BETA_{t-1}$ = market-model beta calculated from the daily stock returns in year $t-1$ with a minimum of 125 observations;
 $IRISK_{t-1}$ = logarithm of the standard deviation of the market-model residuals calculated from the daily stock returns in year $t-1$ with a minimum of 125 observations;
 ΔLEV_t = change in leverage from year $t-1$ to year t where LEV is the leverage ratio, which is defined as the ratio of total debt (DLTT + DLC) divided by total assets;
 $\Delta DP_t = DP_t - DP_{t-1}$; where DP is the ratio of dividends (DVC) to the market value of equity;

(continued on next page)

$\Delta EP_t = EP_t - EP_{t-1}$; where EP is the ratio of income before extraordinary items (IB) to the market value of equity;
 $\Delta MB_t = MB_t - MB_{t-1}$; where MB is the market-to-book value of equity, which is defined as the market value of equity divided by the book value of equity (CEQ);
 $\Delta SGR_t = SGR_t - SGR_{t-1}$; where SGR is the percentage change in annual sales (SALE);
 $\Delta RATE_t = RATE_t - RATE_{t-1}$; where $RATE$ is the S&P stock rating (9 = A+, 8 = A, 7 = A-, 6 = B+, 5 = B, 4 = B-, 3 = C, 2 = D, 1 = not rated); and
 $\Delta SHRS_t = SHRS_t - SHRS_{t-1}$; where $SHRS$ is the logarithm of shares outstanding.

$$\begin{aligned}\Delta COVERAGE_{i,t+1} = & \beta_0 + \beta_1 DISCI_{i,t} + \beta_2 HIPERFORM_{i,t} + \beta_3 DISCI_{i,t} * HIPERFORM_{i,t} \\ & + \beta_4 \Delta SIZE_{i,t} + \beta_5 \Delta STDROE_{i,t} + \beta_6 \Delta INVPRICE_{i,t} + \beta_7 \Delta RETVAR_{i,t} \\ & + \beta_8 \Delta RD_{i,t} + \beta_9 \Delta ROA_{i,t} + \beta_{10} \Delta CORR_{i,t} + \varepsilon_{i,t},\end{aligned}\quad (5)$$

$$\begin{aligned}\Delta |FE|_{i,t+1} = & \beta_0 + \beta_1 DISCI_{i,t} + \beta_2 HIPERFORM_{i,t} + \beta_3 DISCI_{i,t} * HIPERFORM_{i,t} + \beta_4 \Delta SIZE_{i,t} \\ & + \beta_5 \Delta STDROE_{i,t} + \beta_6 ACHEPS_{i,t} + \beta_7 \Delta RD_{i,t} + \beta_8 \Delta ROA_{i,t} + \beta_9 \Delta CORR_{i,t} + \varepsilon_{i,t},\end{aligned}\quad (6)$$

$$\begin{aligned}\Delta DISP_{i,t+1} = & \beta_0 + \beta_1 DISCI_{i,t} + \beta_2 HIPERFORM_{i,t} + \beta_3 DISCI_{i,t} * HIPERFORM_{i,t} + \beta_4 \Delta SIZE_{i,t} \\ & + \beta_5 \Delta STDROE_{i,t} + \beta_6 ACHEPS_{i,t} + \beta_7 \Delta RD_{i,t} + \beta_8 \Delta ROA_{i,t} + \beta_9 \Delta CORR_{i,t} + \varepsilon_{i,t},\end{aligned}\quad (7)$$

where *COVERAGE* is the 12-month average of the number of analysts who issue annual earnings forecasts captured in the I/B/E/S database for a specific firm; *|FE|* is the absolute value of the 12-month average of analyst forecast errors, which is defined as actual earnings minus the mean forecast, deflated by the stock price at the beginning of the fiscal year; and *DISP* is the 12-month average of the standard deviations of analyst forecasts, deflated by the stock price at the beginning of the fiscal year.

We include a number of control variables derived from prior research. We include firm size (*SIZE*) because larger firms have more potential brokerage or investment banking businesses for analysts' brokerage houses (Bhushan 1989), which affects analyst forecasting behavior. We include the inverse of stock prices (*INVPRICE*) because Brennan and Hughes (1991) suggest that it proxies for the brokerage commission rate. Analysts are more likely to follow firms with higher levels of return variability because the anticipated trading benefits based on private information on these stocks are greater (Bhushan 1989). We therefore include *STDROE*, which is measured as the standard deviation of *ROE* in the preceding four quarters, and *RETVAR*, computed as the daily stock return variance over the 200 days prior to the year-end. We include research and development expense (*RD*) as a proxy for the level of information asymmetry (Aboody and Lev 2000) because analysts have relatively stronger incentives to follow firms with higher levels of information asymmetry (Barth et al. 2001). The earnings-return (Pearson) correlation (*CORR*) between *ROE* and annual stock returns in the preceding four quarters captures the difficulty in predicting a firm's earnings. In addition, *ROA* controls for firm profitability. Finally, annual change in *EPS* (*ACHEPS*) controls for the magnitude of the forthcoming earnings information (Ali et al. 2007). All other variables are as defined earlier.

Table 6, Panel A presents a comparison of the levels of and changes in the three main analyst variables in the year following first-time CSR disclosures. Initiators are covered by more analysts than non-initiators (*COVERAGE*: 26.08 for initiators versus 15.72 for non-initiators, $p < 0.01$), and achieve greater improvement in forecast accuracy than non-initiators ($\Delta |FE|$: -0.137 for initiators versus 0.120 for non-initiators, though at a more marginal statistical significance level with $p = 0.07$).

We present the multivariate regression results for Equations (5), Equation (6), and Equation (7) in Panels B, C, and D of Table 6, respectively. Column I of Panel B shows that there is a significantly positive coefficient on *DISCI * HIPERFORM* (coeff. = 1.052, $p = 0.05$), which suggests that analyst following increases for initiators with superior CSR performance. When we run the regression separately in the two subsamples portioned based on industry medians of CSR

TABLE 6
Post-CSR Disclosure Changes in Analyst Forecasts

Panel A: Mean Comparison

	Full Sample	$DISCI_t = 1$	$DISCI_t = 0$	t-value (difference)
$COVERAGE_{t+1}$	15.768	26.079	15.718	7.95
$ FE _{t+1}$	0.226	0.215	0.226	-0.22
$DISP_{t+1}$	0.205	0.237	0.205	1.08
$\Delta COVERAGE_{t+1}$	-1.250	-0.753	-1.252	-1.04
$\Delta FE _{t+1}$	0.011	-0.137	0.120	-1.85
$\Delta DISP_{t+1}$	0.009	-0.011	0.009	-0.70

Panel B: Post-CSR Disclosure Analyst Coverage (Dependent Variable = $\Delta COVERAGE_{t+1}$)

Variables	I		II		III	
	Full Sample		Top 50% PERFORM		Bottom 50% PERFORM	
Intercept	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.
$DISCI_t$	-0.409***	-16.88	-0.401***	-8.94	-0.506***	-21.86
$HIPERFORM_t$	-0.439	-1.01	0.904**	2.12	0.008	0.03
$DISCI_t * HIPERFORM_t$	-0.020	-0.50				
$\Delta SIZE_t$	1.052**	1.95				
$\Delta STDROE_t$	0.493***	8.45	0.855***	5.59	0.471***	7.71
$\Delta INVPRICE_t$	-0.112	-0.93	-0.201	-0.68	-0.076	-0.60
$\Delta RETVAR_t$	4.201***	8.44	9.552***	6.53	3.229***	6.41
ΔRD_t	0.008	0.83	-0.009	-0.44	0.027***	2.46
ΔROA_t	3.660***	7.89	0.144	0.10	4.056***	8.65
$\Delta CORR_t$	1.751***	7.11	2.189***	3.60	1.616***	6.18
Adjusted R ²	0.001	0.89	0.001	0.69	0.001	1.21
	0.026		0.016			0.017

(continued on next page)

Panel B: Post-CSR Disclosure Analyst Coverage (Dependent Variable = $\Delta \text{COVERAGE}_{i,t+1}$)

Variables	I		II		III	
	Full Sample		Top 50% PERFORM		Bottom 50% PERFORM	
	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.
n ($\text{DISCI}_t = 1$)		213		165		48
n		15,298		7,664		7,634

Panel C: Post-CSR Disclosure Analyst Forecast Errors (Dependent Variable = $\Delta |\text{FE}|_{i,t+1}$)

Variables	I		II		III	
	Full Sample		Top 50% PERFORM		Bottom 50% PERFORM	
	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.
Intercept		76.49		54.07		66.68
DISCI_t	1.134***	-1.92	0.923***	-1.77	1.124***	-1.42
HIPERFORM_t	-0.496**	-5.41	-0.251*		-0.425	
$\text{DISCI}_t * \text{HIPERFORM}_t$	-0.261***	1.04				
ΔSIZE_t	0.318	-10.74	-0.198***	-5.10	-0.293***	-5.56
ΔSTDROE_t	-0.277***	1.71	0.059	0.52	0.189*	1.89
ΔACHEPS_t	0.129*	-2.59	-0.065	-0.85	-0.140**	-2.09
ΔRD_t	-0.131***	-1.66	-1.513***	-3.83	0.250	0.79
ΔROA_t	-0.410*	7.46	0.187	1.28	0.855***	8.07
ΔCORR_t	0.610***	-0.29	0.001	0.32	-0.001	-0.36
Adjusted R ²	-0.001					
n ($\text{DISCI} = 1$)	0.020		0.011		0.016	
n	213		165		48	
	13,186		6,587		6,599	

(continued on next page)

Panel D: Post-CSR Disclosure Analyst Forecast Dispersion (Dependent Variable = $\Delta DISP_{t+1}$)

Variables	I		II		III	
	Full Sample		Top 50% PERFORM		Bottom 50% PERFORM	
	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.
Intercept	0.007***	5.59	0.006***	4.67	0.007***	4.98
$DISC_{it}$	0.008	0.37	-0.048***	-3.21	-0.020	-1.05
$HIPERFORM_{it}$	0.001	0.24				
$DISC_{it} * HIPERFORM_{it}$	-0.053**	-2.08				
$\Delta SIZE_{it}$	-0.015***	-6.67	-0.008**	-2.40	-0.021***	-6.62
$\Delta STDROE_{it}$	0.002	0.33	-0.005	-0.62	0.010	1.09
$\Delta ACHEPS_{it}$	-0.064***	-6.34	-0.100***	-7.12	-0.031**	-2.11
ΔRD_{it}	-0.022	-1.04	0.057*	1.76	-0.071**	-2.38
ΔROA_{it}	-0.051***	-4.10	-0.058***	-3.44	-0.053***	-2.81
$\Delta CORR_{it}$	-0.001	-0.49	0.001	1.67	-0.001	-2.01
Adjusted R ²	0.024		0.011		0.011	
n ($DISC_{it} = 1$)	213		165		48	
n	14,363		7,153		7,210	

*, **, *** Indicate that the estimated coefficient is statistically significant at the 10 percent, 5 percent and 1 percent levels, respectively.

All continuous variables are winsorized at the 1st and 99th percentiles. All t-statistics are corrected using the Huber-White procedure.

Variable Definitions:

$DISC_{it}$ = indicator variable that equals 1 if it is the earliest reporting year of a firm that issues CSR reports (CSR report initiating year), and 0 otherwise;

$HIPERFORM_{it}$ = indicator variable that equals 1 if the firm's total CSR strength score, namely, $PERFORM_{it}$, is higher than the industry median (in other words, if the firm is classified as a high CSR performer in its industry), and 0 otherwise;

$\Delta COVERAGE_{t+1} = COVERAGE_{t+1} - COVERAGE_t$, where $COVERAGE_t$ is the 12-month average of the number of analysts who issued annual earnings forecasts in $IB/E/S$ for the firm of interest;

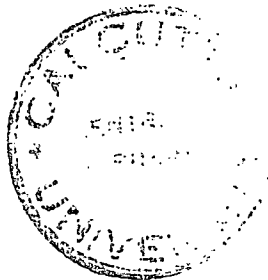
$\Delta |FE|_{t+1} = |FE|_{t+1} - |FE|_t$, where $|FE|_t$ is the absolute value of the 12-month average of analyst forecast errors defined as actual earnings minus the mean forecast, deflated by the stock price at the beginning of the fiscal year;

$\Delta DISP_{t+1} = DISP_{t+1} - DISP_t$, where $DISP_t$ is the 12-month average of the standard deviation of analyst forecasts, deflated by the stock price at the beginning of the fiscal year;

$SIZE_t$ = natural logarithm of the market value of equity ($CSHO_t * PRCC_{t,F}$) at the beginning of the year;

(continued on next page)

STDROE = standard deviation of ROE (IB/CEQ) in the four quarters of the preceding year;
INVPRICE = inverse of the stock price (PRCC_F) at the beginning of the fiscal year;
RETVAR = daily stock return variance estimated over the 200 days prior to the year-end;
RD = research and development expense (XRD) deflated by total assets at the beginning of the fiscal year;
ROA = earnings before extraordinary items divided by total assets; and
CORR = Pearson correlation coefficient between ROE and annual stock returns in the four quarters of the preceding year.



performance, we find a significantly positive coefficient on *DISCI* (coeff. = 0.904, $p = 0.03$) but only in the better performing group.²¹

We obtain similar results for the absolute magnitude of analyst forecast errors. Table 6, Panel C demonstrates that, while in the full sample, the coefficient on *DISCI * HIPERFORM* is insignificant, the main effect of *DISCI* in the partitioned sample regression is marginally negative (coeff. = -0.251, $p = 0.08$) for the better performing group. The results for forecast dispersion, presented in Table 6, Panel D show a similar pattern. In the full-sample regression in Column I, the coefficient on *DISCI * HIPERFORM* is significantly negative (coeff. = -0.053, $p = 0.04$). The regressions in the partitioned sample (Columns II and III) yield a significantly negative coefficient on *DISCI* only among better performing firms (coeff. = -0.048, $p < 0.01$).

In sum, voluntary CSR disclosure is associated with increased analyst coverage, improved forecast accuracy, and a reduction in forecast dispersion among firms with relatively superior CSR performance. These results are consistent with our conjecture that CSR disclosure by strong CSR performers helps reduce information asymmetry between managers and shareholders and among shareholders. The evidence supports our reasoning that CSR disclosure can reduce the cost of equity capital by reducing estimation risk in the market. In other words, consistent with our earlier evidence that the cost of equity capital benefit manifests only among firms with relatively superior CSR performance, the effects of CSR disclosure on analyst coverage and forecast accuracy and dispersion are significant only when firms achieve relatively superior CSR performance.

CSR Disclosure and Subsequent Equity Issuances

As discussed previously, we predict that firms anticipating external financing needs are more likely to initiate CSR disclosures in the hope of obtaining cheaper capital. Hence, we should observe more equity issuances after first-time CSR disclosures. We estimate the following logistic regression for Equation (8) and OLS regression for Equation (9) to empirically test this prediction:

$$\begin{aligned} \log[\text{prob}(\text{SEO}_{i,t+T})/(1 - \text{prob}(\text{SEO}_{i,t+T}))] = & \beta_0 + \beta_1 \text{DISCI}_{i,t} + \beta_2 \text{MB}_{i,t} + \beta_3 \text{LNSALES}_{i,t} \\ & + \beta_4 \text{ROA}_{i,t} + \beta_5 \text{LEV}_{i,t} + \beta_6 \text{CASH}_{i,t} + \beta_7 \text{FIN}_{i,t} \\ & + \beta_8 \text{PAYOUT}_{i,t} + \beta_9 \text{CAPITAL}_{i,t} + \beta_{10} \text{RD}_{i,t} \\ & + \beta_{11} \text{LNDISP}_{i,t} + \varepsilon_{i,t}, \end{aligned} \quad (8)$$

$$\begin{aligned} \text{ISSUE\$}_{i,t+T} = & \beta_0 + \beta_1 \text{DISCI}_{i,t} + \beta_2 \text{MB}_{i,t} + \beta_3 \text{LNSALES}_{i,t} + \beta_4 \text{ROA}_{i,t} + \beta_5 \text{LEV}_{i,t} + \beta_6 \text{CASH}_{i,t} \\ & + \beta_7 \text{FIN}_{i,t} + \beta_8 \text{PAYOUT}_{i,t} + \beta_9 \text{CAPITAL}_{i,t} + \beta_{10} \text{RD}_{i,t} + \beta_{11} \text{LNDISP}_{i,t} + \varepsilon_{i,t}, \end{aligned} \quad (9)$$

where T ($= 1$ or 2) denotes one or two years following CSR disclosure, and other notations follow those of earlier regression equations. $\text{SEO}_{i,t+1}$ ($\text{SEO}_{i,t+2}$) equals 1 if a firm conducts a seasoned equity offering within one (two) year(s) following CSR disclosure. $\text{ISSUE\$}_{i,t+1}$ ($\text{ISSUE\$}_{i,t+2}$) is the total dollar amount in billions raised in SEOs within one (two) year(s) following CSR disclosure. We obtain information on SEOs from the Security Data Corporation (SDC).

Following prior studies, we control for other potential factors affecting the equity issuance decisions of firms. We include the market-to-book ratio (*MB*), as Stein (1995) suggests that firms will choose the time when their stock is overvalued to issue equity. In addition, growth firms have

²¹ A careful examination of the distribution of $\Delta \text{COVERAGE}$ reveals that there are some relatively large values on the negative tail. To determine whether our results are sensitive to these large values, we exclude the bottom 5 percent of $\Delta \text{COVERAGE}$ and get a more symmetric distribution on this variable. We also try excluding extreme values of the upper and the bottom 2 percent of this variable. Overall, our main results are not sensitive to the treatment of the large negative values of $\Delta \text{COVERAGE}$.

a greater need for capital. We include research and development expenses (*RD*) and capital expenditures (*CAPITAL*) as additional proxies for growth opportunities. The likelihood of an equity issuance can depend on the extent of the financial constraints that a firm faces. Also, firms may follow a specific pecking order in their choice of financing options and rely preferentially on internal reserves and debt financing before issuing equity (Myers and Majluf 1984). Hence, we include profitability (*ROA*), cash flow (*CASH*), payout ratio (*PAYOUT*: cash dividend) and leverage (*LEV*) to capture financial constraints, internally generated funds, and debt capacity. We control for analyst forecast dispersion (*LNDISP*: the logarithm of the standard deviation of analyst forecasts divided by the consensus forecast) as a proxy for the degree of agreement between management and investors because Dittmar and Thakor (2007) argue that firms are more likely to issue equity when the level of agreement is high. Finally, we control for firm size (*LNSALES*: logarithm of total sales) and financing activities (*FIN*) already conducted in the current year.

Equation (8) assesses whether CSR disclosure is related to the likelihood of future equity issuance through SEOs, and Equation (9) examines whether CSR disclosure is associated with the size of SEOs. Table 7, Panel A indicates that firms are more likely to seek equity capital through SEOs in the two years following CSR disclosure. The coefficients on *DISCI* are positive and marginally significant (coeff. = 0.504, $p = 0.09$ for SEO_{t+1} ; coeff. = 0.625, $p = 0.06$ for SEO_{t+2}). Based on the coefficient estimate for *DISCI* in Column I (II), in the first year (two years) after CSR disclosure, the odds that initiators will issue equity is 65.5 percent (86.8 percent) higher than that of non-initiators.

Table 7, Panel B reveals that, holding other factors constant, disclosing firms not only are more likely to issue equity, but also raise a significantly larger amount than non-initiators. The difference ranges from U.S. \$165 million ($ISSUES_{t+2}$) to U.S. \$173 million ($ISSUES_{t+1}$). These results are consistent with managers initiating CSR disclosures before going to the capital market in anticipation of obtaining cheaper external capital and increasing their capacity to raise external capital.

Additional Analyses

Alternative Measures of CSR Disclosure

Our CSR disclosure measure, *DISCI*, which captures first-time reporters, best serves the purpose of testing our hypotheses. However, for the sake of completeness, we also test our hypotheses using alternative measures of CSR disclosure. The first measure, $DISCI_{i,t}$, is an indicator variable that equals 1 if Firm *i* discloses a standalone CSR report in Year *t*, and 0 otherwise. The second measure, $DISCN_{i,t}$, indicates whether a firm only sporadically publishes CSR reports. $DISCI_{i,t}$ takes a value of 1 if Firm *i* issues a CSR report in year *t*, but not in year *t*+1 (even though it resumes reporting in year *t*+2 or later), and 0 otherwise. Allowing for non-first-time disclosing years significantly increases our sample size. We obtain results qualitatively similar to those based on *DISCI*. Finally, after excluding first-time reports, we use only continuing reports, $DISCI_{i,t}$, which can be just mundane duplications of earlier reports containing less incremental information. We find a much weaker, albeit still significant, result.

Individual Measures of the Cost of Equity Capital

In the above analyses, we use the average of the three cost of equity capital measures based on Gebhardt et al. (2001), Claus and Thomas (2001), and Easton (2004). These estimation methods are based on different earnings growth assumptions and therefore have distinct strengths and weaknesses. The merit of each measure is debated among researchers, and it is not our intention to resolve this debate. Averaging across the three estimates potentially reduces noise in individual measures (Larcker and Rusticus 2010), and is widely used in the literature (Hail and Leuz 2006;

TABLE 7 (continued)

$DISC_{it}$	= indicator variable that equals 1 if it is the earliest reporting year of a firm that issues CSR reports (CSR report initiating year), and 0 otherwise;
$HIPERFORM_{it}$	= indicator variable that equals 1 if the firm's total CSR strength score, namely, $PERFORM_{it}$, is higher than the industry median (in other words, if the firm is classified as a high CSR performer in its industry), and 0 otherwise;
$SEO_{t+1} (SEO_{t+2})$	= 1 if a firm conducts a seasoned equity offering within one (two) year(s) after CSR disclosure, and 0 otherwise;
$ISSUEAMT_{t+1} (ISSUEAMT_{t+2})$	= total dollar amount in billions issued by a firm's seasoned equity offering within one (two) year(s) after the firm issues a CSR report;
MB_{it}	= market-to-book ratio, which is defined as the market value of equity ($CSHO * PRCC\_F$) divided by the book value of equity (CEQ);
$LNSALES_{it}$	= natural logarithm of total sales (SALE);
LEV_{it}	= leverage ratio, which is defined as the ratio of total debt ($DLTT + DLC$) divided by total assets;
$CASH_{it}$	= total cash-to-asset ratio, which is defined as cash and short-term investments (CHE) divided by total assets;
FIN_{it}	= amount of debt or equity capital raised by the firm scaled by total assets at the beginning of year t . It is measured as the sale of common stock and preferred shares minus the purchase of common stock and preferred shares ($SSTK - PRSTKC$) plus the long-term debt issuance minus the long-term debt reduction ($DLTIS - DLTR$);
$PAYOUT_{it}$	= cash dividend (DVC) relative to total assets;
$CAPITAL_{it}$	= capital expenditures (CAPX) scaled by total assets;
RD_{it}	= research and development expense (XRD) scaled by total assets. We assume a value of 0 if a firm's research and development expense is missing; and
$LNDISP_{it}$	= forecast dispersion in year t , which is measured as the logarithm of the standard deviation of analyst estimates of year t earnings divided by the consensus forecast of year t earnings.

Dhaliwal et al. 2006). To provide assurance that our results are not sensitive to the choice of these measures, we repeat our analyses using the three measures separately and obtain similar results.

Correlation between KLD Performance and CSR Disclosure

KLD performance scores (*PERFORM*) are correlated with the decision of firms to issue CSR reports for the first time (*DISCI*), but their correlation coefficients (Table 2, Panel B) are moderate at about 10 percent.²² As discussed earlier, we control for CSR performance in all regression equations. To further alleviate the concern regarding the correlation between the KLD performance scores and CSR disclosure of firms and the potential impact of this correlation on our results, we conduct the following robustness analyses. First, we remove the transparency-related category, that is, "Corporate Governance," from the KLD performance ranking scores, as this category contains a subcategory, "Transparency," which is a dimension that is likely to reflect the CSR disclosure policy of firms. Our main inferences are unchanged. Second, we use the performance score of each KLD CSR category to measure firms' social performance. Our main inferences are unchanged. Finally, we match each *DISCI* observation with a non-disclosing firm that has the closest industry-adjusted KLD CSR performance score in the same year and industry and run regression Equation (2). The coefficient on *DISCI* is significantly negative at the conventional level, suggesting that CSR initiators enjoy a subsequent reduction in the cost of equity capital.

Alternative Measures of CSR Performance

To determine if our results are sensitive to alternative measures of CSR performance, we use two additional measures. One measure is an indicator *DJSI* that equals 1 if a firm appeared in the Dow Jones Sustainability Index in any year during the period 2002–2007, and 0 otherwise. The other is an indicator *CRO* that equals 1 if a firm was on the "100 Best Corporate Citizens" list for 2007 from the Corporate Responsibility Officer, and 0 otherwise.²³ We do not consider year-to-year variation because of data constraint. It turns out *DJSI* (*CRO*) is correlated with our KLD performance scores with a Pearson coefficient of 30 percent (23 percent). Using these two measures in place of the KLD scores produces qualitatively similar results.

V. SUMMARY AND CONCLUSIONS

We examine a potential benefit associated with the initiation of voluntary disclosure of CSR activities: a reduction in the cost of equity capital. We find that the likelihood of a firm initiating standalone disclosure of CSR activities is associated with a higher prior year cost of equity capital. Firms with CSR performance superior to that of their industry peers enjoy a reduction in the cost of equity capital after they initiate CSR reports. Further, firms initiating CSR disclosure with superior CSR performance attract dedicated institutional investors and analyst coverage, and these analysts achieve lower absolute forecast errors and dispersion following such disclosure. Finally, CSR disclosure initiators appear to exploit this potential benefit of a reduction in the cost of equity

²² Consistent with this low correlation, KLD provides the following information regarding how it rates the CSR performance of firms (see <http://www.kld.com/research/methodology.html>): "KLD researches the social, environmental, and governance performance of corporations. KLD research relies on five distinct data sources to inform our ratings and analysis. Data are collected in a disciplined process from a wide variety of company, government, and non-government organization and media sources. KLD tracks each company through more than 14,000 global media sources daily." These five distinct data sources include (1) direct communication with company officers; (2) a network of global ESG research firms that cover non-U.S. markets; (3) review of more than 14,000 global news sources; (4) public documents of companies, including annual reports and proxy statements; and (5) information obtained from government and non-government organizations including the U.S. Department of Labor, EPA, Human Rights Watch, OSHA, CANICOR, Ceres, ICCR, and DoD. Hence, it appears that the CSR reports of firms constitute only one of the numerous information sources employed by KLD.

²³ See <http://www.thecro.com/>.

capital. They are more likely than non-disclosing firms to conduct SEOs to raise capital in the two years following the disclosure. In addition, among firms conducting SEOs, CSR disclosure initiators raise a significantly larger amount of equity capital than non-initiators.

This study adds to the voluntary disclosure literature by extending the traditional research on voluntary disclosure beyond the narrow focus of financial disclosure. Our analyses enhance our understanding of the rationales behind and the consequences of the recent trend in voluntary CSR disclosure. These results have important implications for companies, regulators, and investors.

A few caveats are worth noting. Most of the control variables that we use in the CSR determination model are obtained from the standard voluntary disclosure literature. To the extent that CSR disclosure is distinct from other forms of voluntary disclosure examined in the literature, we may have missed important determinants of CSR disclosure. In addition, it is possible that we missed some reports on stale websites because of their lack of maintenance, which would add noise to our results. Also, we do not examine the content of the CSR reports. To the extent that the detailed information of these reports is not fully captured by the KLD scores, we fail to capture some important characteristics of CSR reports. Further, it is important to control for the other disclosure policies of firms when examining the impact of CSR disclosure. Our empirical proxies using management guidance and earnings quality may not be sufficient to capture these potential confounding effects. Finally, although the KLD rating is widely used in the management literature, a significant amount of future research is warranted to further establish its validity in measuring the social performance of firms.

These caveats notwithstanding, we believe that our study opens various venues for future research. For example, CSR disclosure and performance could have a different impact on the cost of debt as debtholders have a payoff function different from that of equityholders. Further, the effect of CSR disclosure could be a function of differences in legal environment and institutional setting. Therefore, an international study could help us better understand CSR disclosure. Last, as mentioned previously, it would be worthwhile to investigate the effect of the information content of CSR reports on the valuation decisions of investors.

APPENDIX CSR CATEGORIES AND PERFORMANCE SCORES BY INDUSTRY

Industries	Total Strength	Community	Corporate Governance	Diversity	Employee Relations	Environment	Human Rights	Product
Mining/Construction	0.789	0.143	0.130	0.189	0.200	0.084	0.027	0.016
Food	2.351	0.399	0.231	1.065	0.401	0.203	0.022	0.028
Textiles/Print/Publish	1.751	0.159	0.143	0.670	0.322	0.296	0.003	0.158
Chemicals	2.224	0.298	0.145	0.713	0.525	0.442	0.000	0.101
Pharmaceuticals	1.599	0.201	0.136	0.752	0.309	0.116	0.000	0.085
Extractive	1.266	0.147	0.159	0.210	0.488	0.241	0.002	0.020
Manf: Rubber/Glass/etc.	1.638	0.311	0.168	0.485	0.404	0.162	0.036	0.072
Manf: Metal	1.186	0.060	0.126	0.204	0.426	0.285	0.000	0.085
Manf: Machinery	1.300	0.134	0.128	0.316	0.318	0.237	0.004	0.163
Manf: Electrical Eqpt	1.397	0.138	0.135	0.426	0.331	0.182	0.007	0.179
Manf: Transport Eqpt	1.985	0.193	0.176	0.606	0.537	0.330	0.000	0.144
Manf: Instruments	1.398	0.127	0.184	0.557	0.230	0.174	0.001	0.126
Manf: Misc.	1.970	0.463	0.232	0.713	0.335	0.055	0.030	0.140
Computers	1.687	0.120	0.183	0.769	0.391	0.114	0.001	0.110
Transportation	1.273	0.118	0.139	0.631	0.269	0.041	0.002	0.076
Utilities	1.919	0.203	0.173	0.721	0.364	0.440	0.001	0.016
Retail: Wholesale	0.757	0.030	0.119	0.341	0.122	0.065	0.000	0.081
Retail: Misc.	1.598	0.192	0.149	0.870	0.230	0.070	0.016	0.071
Retail: Restaurant	1.545	0.110	0.140	0.932	0.235	0.110	0.004	0.015
Financial	1.624	0.352	0.265	0.710	0.227	0.009	0.005	0.055
Insurance/Real Estate	0.534	0.036	0.153	0.256	0.085	0.004	0.000	0.001
Services	0.822	0.026	0.146	0.492	0.106	0.020	0.000	0.032
Mean	1.478	0.180	0.174	0.611	0.292	0.137	0.005	0.078

This table provides a brief summary of the seven categories included in the KLD Research & Analytics database, which are used to rate the CSR performance of firms. The data cover the 1991-2007 period. Within each of these seven categories, KLD defines a set of potential strengths and assigns a value of 1 if the strength exists, and 0 otherwise. The statistics provided in this table are the mean performance scores (non-industry-adjusted raw performance scores) for each industry. The seven main categories and their subcategories are summarized below.

Main Categories	Sub-Categories	Max. Strength (Perfect Score)	Actual Max. Strength	Actual Mean Strength
Community	(1) Charitable Giving, (2) Innovative Giving, (3) Non-U.S. Charitable Giving, (4) Support for Education, (5) Support for Housing, (6) Volunteer Programs, and (7) Other Strengths	7	5	0.189
Corporate Governance	(1) Compensation, (2) Ownership, (3) Political Accountability, (4) Transparency, and (5) Other Strengths	5	3	0.167
Diversity	(1) Board of Directors, (2) CEO, (3) Employment of the Disabled, (4) Promotion, (5) Women and Minority Contracting, (6) Work/Life Benefits, (7) Gay and Lesbian Policies, and (8) Other Strengths	8	7	0.605
Employee Relations	(1) Health and Safety, (2) Retirement Benefits, (3) Union Relations, (4) Cash Profit Sharing, (5) Employee Involvement, and (6) Other Strengths	6	5	0.292
Environment	(1) Beneficial Products and Services, (2) Clean Energy, (3) Pollution Prevention, (4) Recycling, and (5) Other Strengths	5	4	0.140
Human Rights	(1) Labor Rights, (2) Relations with Indigenous Peoples, and (3) Other Strengths	3	2	0.004
Product	(1) Benefits the Economically Disadvantaged, (2) Quality, (3) R&D/Innovation, and (4) Other Strengths	4	3	0.077
Total Strength	The sum of all of the above seven main categories.	38	29	1.474

REFERENCES

- Aboody, D., and B. Lev. 2000. Information asymmetry, R&D, and insider gains. *The Journal of Finance* 55: 2747–2766.
- Ali, A., T. Y. Chen, and S. Radhakrishnan. 2007. Corporate disclosures by family firms. *Journal of Accounting and Economics* 44 (1–2): 238–286.
- Al-Tuwaijri, S. A., T. E. Christensen, and K. E. Hughes, II. 2004. The relations among environmental disclosure, environmental performance, and economic performance: A simultaneous equations approach. *Accounting, Organizations and Society* 29: 447–471.
- Amihud, Y., and H. Mendelson. 1986. Asset pricing and the bid-ask spread. *Journal of Financial Economics*

- 17: 223–249.
- Anderson, J. C., and A. W. Frankel. 1980. Voluntary social reporting: An ISO-beta portfolio analysis. *The Accounting Review* 15: 467–479.
- Baron, D. P., M. A. Harjoto, and H. Jo. 2009. The economics and politics of corporate social performance. Working paper, Stanford University.
- Barth, M. E., W. H. Beaver, and W. R. Landsman. 1998. Relative valuation roles of equity book value and net income as a function of financial health. *Journal of Accounting and Economics* 25: 1–34.
- , R. Kasznik, and M. F. McNichols. 2001. Analyst coverage and intangible assets. *Journal of Accounting Research* 39: 1–34.
- Berman, S. L., A. C. Wicks, S. Kotha, and T. M. Jones. 1999. Does stakeholder orientation matter? The relationship between stakeholder management models and firm financial performance. *Academy of Management Journal* 42: 488–506.
- Bhushan, R. 1989. Firm characteristics and analyst following. *Journal of Accounting and Economics* 11: 255–274.
- Botosan, C. A. 1997. Disclosure level and the cost of equity capital. *The Accounting Review* 72: 323–349.
- , and M. A. Plumlee. 2002. A re-examination of disclosure level and the expected cost of equity capital. *Journal of Accounting Research* 40: 21–40.
- Brennan, M., and P. Hughes. 1991. Stock prices and the supply of information. *The Journal of Finance* 46: 1665–1691.
- Bushee, B. J. 1998. The influence of institutional investors on myopic R&D investment behavior. *The Accounting Review* 73: 305–333.
- 2001. Do institutional investors prefer near-term earnings over long-run value? *Contemporary Accounting Research* 18: 207–246.
- , and C. F. Noe. 2000. Corporate disclosure practices, institutional investors, and stock return volatility. *Journal of Accounting Research* 38: 171–202.
- Chatterji, A. K., D. I. Levine, and M. W. Toffel. 2009. How well do social ratings actually measure corporate social responsibility? <http://onlinelibrary.wiley.com/doi/10.1111/j.1530-9134.2009.00210.x/abstract> 18: 125–169.
- Claus, J., and J. Thomas. 2001. Equity premia as low as three percent? Evidence from analysts' earnings forecasts for domestic and international stock markets. *The Journal of Finance* 56: 1629–1666.
- Core, J. E. 2001. A review of the empirical disclosure literature: Discussion. *Journal of Accounting and Economics* 31: 441–456.
- Cormier, D., and M. Magnan. 2003. Does disclosure matter? *CA Magazine* 136 (4): 43–45.
- Dechow, P. M., R. G. Sloan, and A. P. Sweeney. 1995. Detecting earnings management. *The Accounting Review* 70: 193–226.
- , S. A. Richardson, and I. Tuna. 2003. Why are earnings kinky? An examination of the earnings management explanation. *Review of Accounting Studies* 8: 355–384.
- Del Guercio, D. 1996. The distorting effect of the prudent-man laws of institutional equity investments. *Journal of Financial Economics* 40: 31–62.
- Dhaliwal, D. S., S. Heitzman, and O. Z. Li. 2006. Taxes, leverage, and the cost of equity capital. *Journal of Accounting Research* 44: 691–723.
- , L. Krull, O. Z. Li, and W. Moser. 2005. Dividend taxes and implied cost of equity capital. *Journal of Accounting Research* 43: 675–708.
- Diamond, D., and R. Verrecchia. 1991. Disclosure, liquidity and the cost of equity capital. *The Journal of Finance* 46: 1325–1360.
- Dietrich, J. R., S. J. Kachelmeier, D. N. Kleinmuntz, and T. J. Linsmeier. 2001. Market efficiency, bounded rationality, and supplemental business reporting disclosures. *Journal of Accounting Research* 39: 243–268.
- Dittmar, A., and A. Thakor. 2007. Why do firms issue equity? *The Journal of Finance* 62: 1–54.
- Dye, R. A. 1985. Disclosure of non-proprietary information. *Journal of Accounting Research* 23: 123–145.
- Easton, P. 2004. PE ratios, PEG ratios, and estimating the implied expected rate of return on equity capital. *The Accounting Review* 79: 73–95.
- Fama, E., and K. French. 1992. The cross-section of expected stock returns. *The Journal of Finance* 47:

- 427–465.
- Francis, J., D. Nanda, and P. Olsson. 2008. Voluntary disclosure, earnings quality, and cost of Capital. *Journal of Accounting Research* 46: 53–99.
- Francis, J., D. Philbrick, and K. Schipper. 1994. Shareholder litigation and corporate disclosures. *Journal of Accounting Research* 32: 137–164.
- Frankel, R., M. McNichols, and P. Wilson. 1995. Discretionary disclosure and external financing. *The Accounting Review* 70: 135–150.
- Gebhardt, W., C. Lee, and B. Swaminathan. 2001. Toward an implied cost of capital. *Journal of Accounting Research* 39: 135–176.
- Gode, D., and P. Mohanram. 2003. Inferring the cost of capital using the Ohlson-Juettner model. *Review of Accounting Studies* 8: 399–431.
- Graham, J. R., C. R. Harvey, and S. Rajgopal. 2005. The economic implications of corporate financial reporting. *Journal of Accounting and Economics* 40: 3–73.
- Graves, S. B., and S. A. Waddock. 1994. Institutional owners and corporate social performance. *Academy of Management Journal* 37: 1035–1046.
- Hail, L., and C. Leuz. 2006. International differences in cost of equity: Do legal institutions and securities regulation matter? *Journal of Accounting Research* 44: 485–531.
- Healy, P., and K. Palepu. 2001. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics* 31: 405–440.
- Hobson, J. L., and S. J. Kachelmeier. 2005. Strategic disclosure of risky prospects: A laboratory experiment. *The Accounting Review* 80: 825–846.
- Hughes, J. S., J. Liu, and J. Liu. 2007. Information asymmetry, diversification, and cost of capital. *The Accounting Review* 82: 705–730.
- Ingram, R. W., and K. B. Frazier. 1980. Environmental performance and corporate disclosure. *Journal of Accounting Research* 18: 614–622.
- Jones, J. 1991. Earnings management during import relief investigations. *Journal of Accounting Research* 29: 193–228.
- KPMG. 2008. KPMG international survey of corporate social responsibility reporting 2008. Available at: <http://www.kpmg.com/LU/en/IssuesAndInsights/Articlespublications/Pages/KPMGInternationalSurveyonCorporateResponsibilityReporting2008.aspx>.
- Lambert, R., C. Leuz, and R. E. Verrecchia. 2007. Accounting information, disclosure, and the cost of capital. *Journal of Accounting Research* 45: 385–420.
- Lang, M., and R. Lundholm. 1993. Cross-sectional determinants of analyst ratings of corporate disclosures. *Journal of Accounting Research* 31: 246–271.
- , and ———. 1996. Corporate disclosure policy and analyst behavior. *The Accounting Review* 71: 467–492.
- Larcker, D. F., and T. O. Rusticus. 2010. On the use of instrumental variables in accounting research. *Journal of Accounting and Economics* 49 (3): 186–205.
- Leftwich, R. W., R. L. Watts, and J. L. Zimmerman. 1981. Voluntary corporate disclosure: The case of interim reporting. *Journal of Accounting Research* 19: 50–77.
- Leuz, C., and C. Schrand. 2008. Voluntary disclosure and the cost of capital: Evidence from firms' responses to the Enron shock. Working paper, University of Chicago.
- , and R. Verrecchia. 2000. The economic consequences of increased disclosure. *Journal of Accounting Research* 38: 91–124.
- , and P. Wysocki. 2008. Economic consequences of financial reporting and disclosure regulation: A review and suggestions for future research. Working paper, University of Chicago.
- Lev, B., C. Petrovits, and S. Radhakrishnan. 2010. Is doing good good for you? How corporate charitable contributions enhance revenue growth. *Strategic Management Journal* 31 (2): 182–200.
- Margolis, J. D., and J. Walsh. 2001. *People and Profits? The Search for a Link between a Firm's Social and Financial Performance*. Mahwah, NJ: Lawrence Erlbaum Publishers.
- Matsumoto, D. A. 2002. Management's incentives to avoid negative earnings surprises. *The Accounting Review* 77: 483–514.
- McWilliams, A., and D. Siegel. 2001. Corporate social responsibility: A theory of the firm perspective.

- Academy of Management Review* 26: 117–127.
- Merton, R. C. 1987. A simple model of capital market equilibrium with incomplete information. *The Journal of Finance* 42: 483–510.
- Myers, S. C., and N. S. Majluf. 1984. Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics* 13: 187–221.
- Orlitzky, M., F. Schmidt, and S. Rynes. 2003. Corporate social and financial performance: A meta-analysis. *Organization Studies* 24: 403–441.
- Plumlee, M., D. Brown, and S. Marshall. 2008. The impact of voluntary environmental disclosure quality on firm value. Working paper, University of Utah.
- Richardson, S., S. H. Teoh, and P. Wysocki. 2004. The walkdown to beatable analysts' forecasts: The role of equity issuance and inside trading incentives. *Contemporary Accounting Research* 21: 885–924.
- Richardson, A., and M. Welker. 2001. Social disclosure, financial disclosure and the cost of equity capital. *Accounting, Organizations and Society* 26: 597–616.
- Rodriguez, P., D. S. Siegel, A. Hillman, and L. Eden. 2006. Three lenses on the multinational enterprise: Politics, corruption and corporate social responsibility. *Journal of International Business Studies* 37: 733–746.
- Rogers, J., and A. Van Buskirk. 2009. Shareholder litigation and changes in disclosure behavior. *Journal of Accounting and Economics* 47 (1–2): 136–156.
- Shleifer, A., and R. W. Vishny. 1986. Large shareholders and corporate control. *The Journal of Political Economy* 94: 461–488.
- Simnett, R., A. Vanstraelen, and W. F. Chua. 2009. Assurance on sustainability reports: An international comparison. *The Accounting Review* 84: 937–967.
- Skinner, D. J. 1997. Earnings disclosures and stockholder lawsuits. *Journal of Accounting and Economics* 23: 249–282.
- Sletten, E. 2008. The effect of stock price on discretionary disclosure. Working paper, Massachusetts Institute of Technology.
- Social Investment Forum. 2007. 2007 report on socially responsible investing trends in the United States. Washington, D.C.: Social Investment Forum.
- Stein, J. C. 1995. Rational capital budgeting in an irrational world. Working paper, Massachusetts Institute of Technology.
- Verrecchia, R. E. 2001. Essays on disclosure. *Journal of Accounting and Economics* 32: 97–180.
- Waddock, S. 2003. Myths and realities of social investing. *Organization & Environment* 16: 369–380.

Should Investors Follow the Prophets or the Bears? Evidence on the Use of Public Information by Analysts and Short Sellers

Michael S. Drake
The Ohio State University

Lynn Rees
Edward P. Swanson
Texas A&M University



ABSTRACT: We investigate whether short sellers and analysts differ in their use of information that is predictive of future returns. We find that short interest is significantly associated in the expected direction with all 11 variables examined. In contrast, analysts tend to positively recommend stocks with high growth, high accruals, and low book-to-market ratios, despite these variables having a negative association with future returns. We then investigate the profitability of using short interest in trading. We find abnormal returns (1.11 percent per month) from a zero-investment strategy that (1) shorts firms with highly favorable analyst recommendations (buy signal) but high short interest (sell signal), and (2) buys firms with highly unfavorable analyst recommendations (sell signal) but low short interest (buy signal). Short interest, therefore, appears to capture predictive information that can be used by investors in trading against analysts' recommendations to increase returns.

Keywords: *short interest; analyst recommendations; fundamental analysis; arbitrage.*

Data Availability: *Data are available from the sources identified in the study.*

I. INTRODUCTION

Academic research provides extensive evidence that fundamental analysis can be used to earn abnormal returns (e.g., Frankel and Lee 1998; Piotroski 2000; Swanson et al. 2003). This evidence suggests there is a delay in the price discovery process, which has spawned research into the roles that financial analysts play in this delay (Bradshaw et al. 2001; Jegadeesh

The authors gratefully acknowledge the helpful comments of Steven J. Kachelmeier (senior editor) and two anonymous referees, Larry Brown, Ted Christensen, Darren Henderson, Devin Shanthikumar, Kay Stice, David Wood, and workshop participants at Brigham Young University, Texas A&M University, The University of Texas at Arlington, 2009 FARS Midyear Meeting, 2009 AAA Annual Meeting, and Sanford Bernstein 2009 Conference on Controversies in Quantitative Finance.

Editor's note: Accepted by Steven Kachelmeier

Submitted: July 2009
Accepted: May 2010
Published Online: January 2011

et al. 2004, hereafter JKKL; Bradshaw 2004; Barniv et al. 2009). Analysts' recommendations are the end-product from an extensive analysis of information, and they affect market prices by stating a specific course of action that an investor should take (Asquith et al. 2005a; Malmendier and Shanthikumar 2007). JKKL 2004 provide evidence, however, that analysts' recommendations are positively associated with some accounting, valuation, and growth characteristics that have a *negative* association with future returns (e.g., low book-to-market, high sales growth, and high accruals). Analysts' incentives to obtain investment banking business and to generate trading commissions are potential explanations for why they tend to over-recommend these stocks (Lin and McNichols 1998; Barber et al. 2007) and, indirectly, why analysts might contribute to a delay in price discovery.

Our study investigates how short sellers use publicly available information. Short sellers are regarded as particularly sophisticated investors under financial economic theory.¹ Similar to analysts, short sellers invest considerable time and resources in analyzing companies, but they face potentially different incentives. Because short sellers place their own capital at risk, they have strong incentives to fully use predictive information. Research on short sellers' use of fundamental information is limited, but Dechow et al. (2001) find that short sellers use fundamentals-to-valuation ratios to identify stocks that are expected to realize negative future returns. We extend Dechow et al. (2001) by investigating how short sellers utilize 11 items of fundamental information identified by JKKL that are predictive of future returns. We also conduct the JKKL analysis of analyst recommendations for our sample to allow a direct comparison of information use by analysts and short sellers. Our investigation provides further evidence on the role of short sellers in capital markets and their ability to interpret publicly available information. While financial analysts are acknowledged for their information intermediary role, we investigate the possibility that short sellers can also serve as an information intermediary and thereby facilitate the price discovery process.

Following the JKKL analysis of analysts' information use, we rank firms into quintile portfolios based on the consensus analyst recommendation. In a corresponding manner, we rank firms into quintile portfolios by the level of short interest to reflect short seller beliefs about future returns. We then employ ordered logistic regression to examine how analysts and short sellers use the company-specific predictive information. We categorize the information into four general types: accounting, valuation, growth, and momentum. We find that short interest is strongly associated with all 11 information items in the direction consistent with their empirical relation to future returns. This result indicates that short sellers are highly informed about how company-specific information is likely to affect future returns. In contrast, we confirm results in JKKL that analysts tend to positively recommend stocks with high growth, high accruals, and low book-to-market ratios, despite these variables having a negative association with future returns. When we examine recommendation revisions, we find similar evidence.

Our finding that short sellers interpret information properly as it relates to future returns supports an information intermediary role for short sellers. Consistent with this role, the SEC (1999, 3) observes that short sellers can "add to stock pricing efficiency because their transactions *inform the market* of their evaluation of future stock price performance" (emphasis added). Pownall and Simko (2005) provide evidence that short sellers can serve as information intermediaries when analyst following for a firm is low. In this situation, there are limited alternative

¹ Diamond and Verrecchia (1987) argue that only informed traders with strong beliefs that stock prices will fall in the near-term will choose to sell stock short. Their reasoning is based on the notion that the high costs of short selling drives out uninformed traders, so that open short positions reflect trades by more informed investors. Boehmer et al. (2008, 491) comment that short sellers "occupy an exalted place in the pantheon of investors as rational, informed market participants who act to keep prices in line."

sources of guidance. Our tests of information use suggest that short sellers can fill a complementary information intermediary role even when coverage by analysts is extensive.²

To provide further evidence on short sellers' effectiveness as information intermediaries, we test whether short interest provides value-relevant information about future returns *beyond* that provided by analyst recommendations and the 11 predictive variables. Short positions could provide incremental information because short sellers have sources of information not considered in our models or because they adjust the weights they place on items of information as market conditions change. In a regression model explaining future six-month returns adjusted for characteristic-based portfolio returns (Daniel et al. 1997), we find that the coefficient on short interest is negative (as predicted) and statistically significant after controlling for the information in analyst recommendations and the 11 predictive variables. These results indicate that short interest provides incremental information about future abnormal returns that is orthogonal to the information provided by analysts and the 11 predictive variables. In addition, the significant negative coefficient on the consensus analyst recommendation indicates that more favorable recommendations result in *lower* future returns after controlling for open short interest and the 11 predictive variables. If we consider recommendation revisions, then its coefficient is insignificant and the coefficient on short interest remains negative and significant.

Our analyses and the evidence from JKKL indicate that analysts sometimes provide favorable (unfavorable) recommendations for stocks with characteristics that are associated with negative (positive) future equity returns. Since analyst recommendations influence trading decisions and stock prices (Asquith et al. 2005a; Malmendier and Shanthikumar 2007), their recommendations can provide support for stock prices that have temporarily deviated from their fundamental values. We investigate if short interest can be used to identify such stocks. That is, we examine if investors can use short interest together with analyst recommendations to construct a portfolio that is likely to earn abnormal future returns. To our knowledge, ours is the first study to link these two investment signals—both from highly regarded capital market participants—to forecast future long-run returns.

We use a large sample of monthly observations over the period 1994 to 2006 to test alternative trading strategies. We first construct quintile portfolios based on the consensus analyst recommendation and then examine abnormal returns from a trading strategy that invests long (short) in firms comprising the best (worst) recommendation quintile. We find that the hedge return is modestly *negative* at -26 basis points per month for our test period. This result is perhaps surprising given the high esteem placed on financial analysts within the financial community, but it is consistent with our results on information use and the growing literature that questions the investment value of analyst recommendations (Demirakos et al. 2004; Bradshaw 2004; Barniv et al. 2009). We employ the same sorting procedure to form investment portfolios based on levels of short interest, and we obtain a statistically significant hedge return of 56 basis points per month from selling short firms in the highest short interest quintile and buying firms in the lowest quintile. Interestingly, the monthly returns from buying stocks with low short interest (29 basis points) are similar to the returns from selling short stocks with high short interest (27 basis points).

² Note that analysts also provide earnings forecasts, price targets, and narrative discussion that can be informative to investors—possibly more informative than their recommendations.

Concurrent research by Boehmer et al. (2010) also finds that significant abnormal returns can be earned by buying stocks with low short interest.³

We next examine whether abnormal returns can be improved by using information from both analysts and short sellers. Here, we intersect the analyst recommendation and short interest quintiles to produce 25 portfolios (5×5) formed using information from both signals. We find that monthly abnormal returns are insignificant for portfolios containing stocks about which analysts and short sellers strongly concur (e.g., least favorable recommendation and high short interest, or most favorable recommendation and low short interest). In contrast, returns are highly significant for portfolios of stocks in which they strongly conflict, if we trade consistent with the short sellers. Specifically, we find that an investor would obtain an average monthly abnormal return of 111 basis points from a zero-investment strategy that (1) invests long in firms with the worst recommendations (sell signal) but the lowest short interest levels (buy signal), and (2) invests short in firms with the best recommendations (buy signal) but the highest short interest levels (sell signal). The monthly abnormal return from this strategy is statistically significant in each sub-period, ranging from 71 basis points in 2004–2006 to 130 basis points in 1999–2003. Our dual-signal approach, therefore, provides the most investment value during the volatile 1999–2003 sub-period. By comparison, following analyst recommendations would cause investors to experience sizable losses over this sub-period, and the returns to trading on short interest in isolation would be statistically insignificant.

Our study contributes to our understanding of short sellers by (1) documenting their efficient use of a number of predictive variables discussed in the academic literature, (2) showing that open short positions are incrementally useful in predicting future returns after controlling for those predictive variables, and (3) showing that the returns to mimicking the trading of short sellers are much larger when conditioned on conflicting analyst recommendations. Collectively, these findings provide a more complete picture of how short sellers influence equity price formation and should be of interest to academics, investors, and regulators. This topic has assumed considerable importance due to the alleged role of short selling in the dramatic decline in stock prices that began with the 2008 credit crisis (Boehmer et al. 2009). Academics are also likely to be interested in the implications of our empirical results for Miller's (1977) theory that binding short sale constraints cause pessimists to be under-represented in price formation, leading to overvaluation when a strong divergence of opinion exists about a stock. Our evidence shows that short sellers are under-represented in price formation whenever they disagree with analysts, regardless of whether they are the optimists or the pessimists.

Section II describes the selection of our sample. In Section III, we investigate the relation of short interest and analyst recommendations with information that has been shown by prior research to predict future returns. Section IV presents returns from trading strategies that use analyst recommendations, short interest, or both of these signals together. This section also reports the results of several robustness tests. We discuss our results and conclude in Section V.

II. SAMPLE SELECTION AND DESCRIPTIVE STATISTICS

To perform our analysis, we require time-series data on analyst recommendations, open short positions, the 11 predictor variables, and stock returns. We obtain analyst recommendations from

³ This finding complements earlier studies that find that stocks with high short interest have significant negative abnormal returns (e.g., Asquith and Meulbroeck 1995; Asquith et al. 2005b; Desai et al. 2002). Notably, abnormal returns from trading on high short interest are not necessarily indicative of market inefficiency due to limits to arbitrage. In contrast, the finding that stocks with low short interest have significant positive returns is clearly inconsistent with market efficiency since buy-and-hold strategies are not subject to similar constraints (see Boehmer et al. [2010] for further discussion).

the Thompson Financial I/B/E/S Recommendations database. Beginning in late 1993, I/B/E/S provides analyst recommendations for a wide cross-section of firms. I/B/E/S codes recommendations into five ordered categories: strong buy = 1; buy = 2; hold = 3; sell = 4; and strong sell = 5. For analyses using recommendations, we reverse this coding (i.e., strong buy = 5; strong sell = 1) to allow for a more intuitive interpretation of our results. Each month, we calculate the consensus recommendation level (*Rec*) as the mean of all outstanding recommendations issued a maximum of 12 months prior to month-end.⁴ We use only the most recent individual analyst recommendation in the calculation. We also require that the individual analyst recommendations be issued on or before the I/B/E/S consensus recommendation date.⁵

In the first set of analyses, we conduct separate tests of analysts' and short sellers' use of information that is predictive of future returns. We perform these tests using analyst recommendations and short interest as of the last month in each calendar quarter, consistent with JKLL.⁶ We include recommendation revisions in these analyses, given that prior research finds that recommendation revisions might be better indicators of future stock price performance than recommendation levels (Womack 1996; JKLL; Barber et al. 2010). We calculate recommendation revisions as the change in recommendation levels from calendar quarter $t-1$ to quarter t (i.e., consecutive quarters). An increase (decrease) in the consensus recommendation indicates an upgrade (downgrade) in the stock relative to the previous calendar quarter $t-1$.⁷

We obtain short interest data to correspond with our analyst recommendations sample period. The Compustat Monthly Securities Database contains monthly short interest for all firms listed on U.S. exchanges beginning in 2003. For earlier years, we purchased monthly short interest data directly from the NYSE, AMEX, and NASDAQ exchanges, and from an online independent vendor.⁸ The stock exchanges report open short positions using the 15th of each calendar month as the settlement date (or the last business day before the 15th). We scale short interest by the number of shares outstanding as reported by CRSP and label the resulting ratio as *Sratio*, which is standard in the literature (e.g., Dechow et al. 2001; Asquith et al. 2005b; Pownall and Simko 2005).⁹

⁴ This requirement helps alleviate concerns that our results are being unduly influenced by stale recommendations, and it is similar to the measure used by JKLL. Thompson Financial claims that recommendations not updated for 180 days are excluded from the I/B/E/S consensus recommendation (see Thompson Financial 2009,11); however, we are uncertain as to how long this policy has been in place. Given our long sample period, we follow procedures implemented in prior research.

⁵ More specifically, I/B/E/S calculates the consensus recommendation on the Thursday before the third Friday of every month (ranging from the 14th to the 20th day of the month). The requirement of excluding recommendations that are issued after this date results in an average delay of 13.7 days between the time when the consensus recommendation is calculated and the beginning of the returns accumulation period. This methodology serves two purposes. First, short interest data are made publicly available mid-month and therefore, both signals—recommendations and short interest—are obtained at approximately the same time during the month. Second, the delay ensures that investors are given ample time to process and impound in price whatever new information is contained in both signals. Thus, we purposely exclude from our tests any drift in stock prices that occurs due to the public disclosure of the signal. In this manner, our methodology differs markedly from the daily rebalancing requirements employed in papers such as Barber et al. (2001) and Barber et al. (2010).

⁶ Performing these analyses using quarterly data is intuitive given that the majority of the predictor variables (seven of the 11) change on a quarterly basis as financial information is disclosed. We find that all inferences are the same when we perform the analyses using monthly data.

⁷ Ljungqvist et al. (2009) provide evidence that the I/B/E/S recommendations database contains systematic errors in the pre-2007 files that is likely to overstate the investment value of analysts' recommendations. Our study is among the first to re-examine the investment value of analysts' recommendations using the cleaned 2007 database.

⁸ Data from the online vendor, shortsqueeze.com, provide less than 1 percent of the total observations. The data cover a period in which we were unable to obtain short interest directly from the NASDAQ. We compared shortsqueeze.com data to that from a six-month period for which we already had short interest data from NASDAQ. The only differences were due to shortsqueeze.com rounding their data to the nearest hundredth place.

⁹ In the "Robustness Tests" section, we examine the sensitivity of our results to deflating by lagged trading volume.

Similar to JKKL, we select a set of 11 predictor variables for our analysis and winsorize each of the predictor variables at the 2.5 and 97.5 percentiles to control for outliers. We group the predictor variables into one of four classifications based on the nature of the variable (see the Appendix for details on the calculation of each variable). The first group, labeled *Accounting*, consists of earnings surprise (*SUE*), total accruals (*TACCR*), and capital expenditures (*CAPEX*). The *Valuation* group consists of the market-value-of-equity (*MVE*), earnings-to-price ratio (*EP*), book-to-market ratio (*BTM*), and the average daily stock turnover (*TURN*). The *Growth* group consists of realized sales growth (*SG*) and forecasted long-term growth (*LTG*). The fourth group, *Momentum*, consists of earnings forecast revision (*FREV*) and price momentum (*MOM*). These variables have been shown in prior research to be associated with future returns (see the Appendix for specific citations). Thus, we expect that sophisticated capital market participants, such as analysts and short sellers, would use information embedded in these variables when establishing their positions.

Finally, we obtain monthly returns data from CRSP to compute future six-month buy-and-hold abnormal returns (*ARET6*) using characteristic portfolio-matching.¹⁰ Specifically, we define abnormal returns as the raw buy-and-hold return adjusted for the portfolio return from 125 benchmark portfolios formed based on size, book-to-market, and momentum ($5 \times 5 \times 5$), as described in Daniel et al. (1997). We use a holding period of six months for consistency with JKKL, but test the sensitivity of our results to alternative holding periods as a supplemental analysis.

Our final sample, resulting from the intersection of Compustat, CRSP, I/B/E/S, and our short interest database, consists of 80,674 firm-quarter observations over the 52 calendar quarters from 1994 to 2006. For our main analyses, we rank firms into quintile portfolios based on analyst recommendations (both levels and changes) and short interest in each calendar quarter *t*. Thus, we rebalance the portfolios quarterly. For recommendation changes, we ensure that firms without a recommendation revision are included in the middle quintile.

Table 1, Panel A presents descriptive statistics for the variables used in our analyses. The mean (median) value for *Rec* of 3.76 (3.79) indicates that the average analyst recommendation is only moderately less than a “buy” (which would be coded 4). A narrow interquartile range of 0.34 (−0.20 to 0.14) for the consensus recommendation change, *ChgRec*, shows that analyst recommendations are generally sticky. Nevertheless, the minimum and maximum values for *ChgRec* indicate that analysts occasionally downgrade a stock all the way from strong buy to strong sell, and *vice versa*. The mean short interest ratio, *Sratio*, is 3.2 percent, which is considerably larger than the median of 1.8 (due to some large values, as indicated by the maximum of 23.5 percent). The mean six-month abnormal return (*Aret6*) is 1.0 percent, but the median is only −1.9 percent.

With respect to the 11 predictor variables, we find that earnings surprise (*SUE*) has a mean of zero but a slightly positive median of 0.002, consistent with most firms reporting earnings that meet or beat the current analyst forecast. Total accruals (*TACCR*) are negative, on average, due to deducting depreciation. Capital expenditures (*CAPEX*) average approximately 6.3 percent of assets. We find that firm size (*MVE*) is highly skewed, with a mean of \$3,578 compared to a median of \$775 (in millions). The average earnings-to-price ratio (*EP*) is only 2.0 percent due to some negative values (median 4.3 percent). The book-to-market ratio (*BTM*) has a mean of 0.50 (median 0.42), consistent with prior research. Approximately 0.58 percent of a firm’s shares turn over on any given day (*TURN*). Realized sales growth (*SG*) averages 17 percent, and analysts’ long-term earnings growth forecasts (*LTG*) average 17.55 percent. Analysts’ forecast revisions (*FREV*) have

¹⁰ If a firm delists during the return accumulation period, we compound the delisting return with the buy-and-hold return and assume the liquidating proceeds are reinvested in a portfolio that earns a normal return for the remainder of the period.

TABLE 1

Descriptive Statistics

Panel A: Summary Statistics for Dependent and Explanatory Variables

Variable	Mean	Std. Dev.	Min	Q1	Median	Q3	Max
<i>Rec</i>	3.76	0.62	1.00	3.33	3.79	4.17	5.00
<i>ChgRec</i>	-0.03	0.42	-4.00	-0.20	0.00	0.14	4.00
<i>SRatio</i>	3.2%	3.8%	0.0%	0.6%	1.8%	4.2%	23.5%
<i>Aret6</i>	0.010	0.371	-1.000	-0.177	-0.019	0.144	12.525
<i>SUE</i>	0.000	0.029	-0.237	-0.004	0.002	0.006	0.312
<i>TACCR</i>	-0.012	0.035	-0.151	-0.028	-0.011	0.005	0.092
<i>CAPEX</i>	0.063	0.061	0.000	0.022	0.045	0.082	0.341
<i>MVE</i>	3,578	8,421	26	255	775	2,512	65,417
<i>EP</i>	0.020	0.118	-1.553	0.016	0.043	0.065	0.214
<i>BTM</i>	0.500	0.343	0.045	0.262	0.424	0.648	2.647
<i>TURN</i>	0.582	0.273	0.000	0.364	0.614	0.819	1.000
<i>SG</i>	1.17	0.26	0.47	1.02	1.11	1.25	2.35
<i>LTG</i>	17.55	8.02	4.25	12.00	15.44	21.07	55.00
<i>FREV</i>	0.000	0.029	-0.163	-0.005	0.002	0.010	0.219
<i>MOM</i>	0.084	0.421	-0.995	-0.125	0.047	0.228	27.827

Panel B: Analyst Recommendations and Short Interest Values by Calendar Year

Year	n	Mean Rec	Mean SRatio	Returns for S&P 500
1994	4,319	3.74	1.6%	1.3%
1995	4,586	3.76	1.7%	37.6%
1996	5,046	3.83	1.8%	23.0%
1997	5,554	3.89	2.2%	33.4%
1998	6,021	3.90	2.3%	28.6%
1999	6,514	3.93	1.9%	21.0%
2000	6,099	4.00	2.0%	-9.1%
2001	5,772	3.87	2.9%	-11.9%
2002	6,559	3.73	3.6%	-22.1%
2003	7,403	3.52	4.0%	28.7%

(continued on next page)

Panel B: Analyst Recommendations and Short Interest Values by Calendar Year

Year	n	Mean Rec	Mean SRatio	Returns for S&P 500
2004	7,686	3.65	4.0%	10.9%
2005	7,663	3.61	4.6%	4.9%
2006	7,452	3.63	5.7%	15.8%
Full Sample	80,674	3.76	3.2%	12.5%

The sample consists of 80,674 firm-quarter observations during the period 1994–2006. See the Appendix for a more detailed description of how each variable is calculated.

Variable Definitions:

- Rec* = consensus analyst recommendation in the last month of the calendar quarter, where 5 = strong buy, 4 = buy, 3 = hold, 2 = sell, and 1 = strong sell. We calculate *Rec* as the mean of all outstanding recommendations issued on or before the I/B/E/S consensus recommendation date; most recent individual analyst recommendation from the previous quarter;
- ChgRec* = change in the consensus analyst recommendation from the previous quarter;
- SRatio* = change in the number of shares sold short as reported for the last month of the calendar quarter divided by the number of shares outstanding as of the same date;
- Aret6* = six-month abnormal return (adjusted for firm size, book-to-market ratio, and past returns) beginning the first day of quarter $t+1$;
- SUE* = seasonally adjusted earnings change scaled by price for fiscal quarter q ;
- TACCR* = total accruals scaled by average assets measured at the end of fiscal quarter q ;
- CAPEX* = rolling sum of the preceding four quarters of capital expenditures ending at fiscal quarter q divided by total assets;
- MVE* = market value of equity at the end of fiscal quarter q ;
- EP* = ratio of the rolling sum of earnings over the preceding four quarters to price at the end of fiscal quarter q ;
- BTM* = ratio of book value of equity to market value of equity as of the end of fiscal quarter q ;
- TURN* = average daily volume per share over the preceding six months;
- SG* = rolling sum of sales growth over the preceding four fiscal quarters;
- LTG* = consensus long-term earnings growth forecast at the end of calendar quarter t ;
- FREV* = rolling sum of the preceding six-month earnings forecast revisions scaled by price; and
- MOM* = price momentum, measured as the six-month raw return ending one month prior to the end of the fiscal quarter q .

a mean of zero but a slightly positive median of 0.002. Price momentum (*MOM*) averages 8.4 percent for the preceding six months (median 4.7 percent).

Table 1, Panel B reports mean analyst recommendations, short interest, and market returns (using the S&P 500 portfolio) for each year from 1994 to 2006. From 1994 to 2000, we observe a monotonic increase in the average analyst recommendation, which peaks at 4 in 2000. The average recommendation then declines in years 2001 through 2003, and remains at a lower level through the end of our test period. This shift corresponds with criticism of analysts that led to the Global Research Analysts Settlement, NASD 2711, and NYSE Rule 472. One line of criticism focused on analysts' conflicts of interest, including their incentive to maintain a positive relation with corporate managers in order to generate investment banking business and to obtain earnings guidance.

Table 1, Panel B also reports another noteworthy change over our test period. The mean level of short interest is around 2 percent from 1994 to 2000. The level then increases appreciably over the next six years, reaching 5.7 percent in the final year of our sample period. This shift, which corresponds with a dramatic increase in the number of hedge funds, increases the importance of research that furthers an understanding of the role of short selling in the price formation process. Note that shifts over time have a minimal effect on our results because we rank firms into quintiles based on their relative values at a given point in time.

III. PREDICTIVE INFORMATION USED BY ANALYSTS AND SHORT SELLERS

In this section, we first examine whether analyst recommendations and short interest incorporate fundamental and other information in the manner shown by prior research to be predictive of future returns. We then investigate whether analyst recommendations and short interest provide information that is incremental to that information.

Univariate Evidence

Table 2 presents mean values for each of the 11 predictive variables by quintile for recommendation levels, recommendation changes, and short interest levels. The quintiles correspond to portfolios that we later use in trading analyses. In Panel A, as we move down each column from the worst to the best recommendations, we find a monotonic (or near monotonic) increase for eight of the 11 variables. The increase for *SUE*, *EP*, *FREV*, and *MOM* is consistent with analyst recommendations properly incorporating the relation of these measures with future returns. In contrast, the increase for *TACCR*, *CAPEX*, *SG*, and *LTG* indicate that analysts misuse this information, which could cause more favorable recommendations to portend lower investment returns. The overall pattern of information use indicates that analysts tend to issue more favorable recommendations for glamour stocks, even though prior studies show that these stocks earn lower subsequent returns (Lakonishok et al. 1994; La Porta 1996; Sloan 1996; Beneish et al. 2001). Examining changes in recommendations, Panel B shows a clear pattern for only three variables; but in each case, the change is consistent with the relation of the information with future returns established in prior research. Specifically, as we move down the columns from downgrades to upgrades, we observe a monotonic increase for earnings forecast revisions (*FREV*) and stock price momentum (*MOM*), and a monotonic decrease for long-term growth (*LTG*). While prior research has generally found that recommendation revisions are better predictors of future returns than are recommendation levels, this analysis indicates that recommendation revisions fail to incorporate eight of the 11 items of predictive information. These results for recommendation levels and changes are similar to the results documented in JKKL.

Panel C of Table 2 provides the corresponding analysis for short interest quintiles. The book-to-market variable (*BTM*) decreases monotonically as the level of short interest increases, so short sellers tend to take greater positions in firms with a higher market value relative to their book

TABLE 2
Mean Values by Quintile for 11 Variables Associated with Future Returns

Panel A: Recommendation Levels														
<i>QRec</i>	<i>n</i>	<i>Rec</i>	<i>SUE</i>	<i>TACCR</i>	<i>CAPEX</i>	<i>MVE</i>	<i>EP</i>	<i>BTM</i>	<i>TURN</i>	<i>SG</i>	<i>LTG</i>	<i>FREV</i>	<i>MOM</i>	
Worst	0.00	16,216	2.92	-0.005	-0.015	0.057	1,923	-0.010	0.659	0.523	1.082	15.16	-0.008	-0.014
	0.25	15,857	3.45	-0.002	-0.013	0.063	4,185	0.020	0.512	0.595	1.134	16.30	-0.002	0.041
	0.50	17,119	3.80	0.001	-0.012	0.065	4,716	0.029	0.468	0.586	1.164	17.42	0.001	0.088
	0.75	15,704	4.07	0.003	-0.010	0.065	4,763	0.031	0.417	0.621	1.212	18.75	0.004	0.139
Best	1.00	15,778	4.59	0.003	-0.009	0.065	2,256	0.031	0.443	0.584	1.237	20.20	0.006	0.171
Panel B: Recommendation Changes														
<i>QChgRec</i>	<i>n</i>	<i>ChgRec</i>	<i>SUE</i>	<i>TACCR</i>	<i>CAPEX</i>	<i>MVE</i>	<i>EP</i>	<i>BTM</i>	<i>TURN</i>	<i>SG</i>	<i>LTG</i>	<i>FREV</i>	<i>MOM</i>	
Down	0.00	16,113	-0.18	0.000	-0.011	0.064	2,711	0.025	0.503	0.597	1.184	17.80	-0.004	0.026
	0.25	13,869	-0.05	0.000	-0.013	0.068	5,633	0.021	0.445	0.653	1.188	17.75	0.000	0.062
No	0.50	22,799	0.00	-0.001	-0.012	0.059	1,903	0.014	0.574	0.497	1.142	17.54	0.000	0.072
	0.75	11,690	0.04	0.001	-0.012	0.066	6,328	0.025	0.421	0.651	1.182	17.46	0.003	0.130
Up	1.00	16,070	0.15	0.001	-0.012	0.063	3,059	0.021	0.499	0.574	1.148	17.20	0.003	0.147
Panel C: Short Interest Levels														
<i>QSlratio</i>	<i>n</i>	<i>Slratio</i>	<i>SUE</i>	<i>TACCR</i>	<i>CAPEX</i>	<i>MVE</i>	<i>EP</i>	<i>BTM</i>	<i>TURN</i>	<i>SG</i>	<i>LTG</i>	<i>FREV</i>	<i>MOM</i>	
Low	0.00	16,118	0.3%	0.000	-0.012	0.059	3,643	0.022	0.622	0.346	1.120	16.10	0.001	0.065
	0.25	16,144	1.1%	0.001	-0.012	0.059	5,287	0.028	0.501	0.492	1.129	15.77	0.001	0.083
	0.50	16,145	2.0%	0.000	-0.012	0.061	4,156	0.025	0.474	0.588	1.142	16.63	0.001	0.096
	0.75	16,144	3.5%	0.000	-0.012	0.065	2,972	0.018	0.459	0.686	1.177	18.28	0.000	0.094
High	1.00	16,123	8.9%	-0.001	-0.011	0.072	1,832	0.009	0.446	0.796	1.259	20.97	-0.002	0.083

The sample consists of 80,674 firm-quarter observations during the period 1994–2006. See Table 1 for descriptions of each variable, and the Appendix for detailed explanations of how each variable is calculated.

value. Three variables increase monotonically as the level of short interest increases: capital expenditures (*CAPEX*), stock turnover (*TURN*), and sales growth (*SG*). In each case, the pattern of information use is consistent with their relation to future returns documented in prior research. Comparing the high and low short interest quintiles for the other variables shows that short positions are generally consistent with how the variables map into future returns, but the relation is not monotonic.

Multivariate Evidence

We next use ordered logistical regression analysis to provide a multivariate test of the relation between analyst and short seller investment signals and the 11 predictor variables. In all regression analyses, we assess statistical significance using test statistics based on standard errors that are adjusted for two-way clustering of residuals by firm and calendar month (Petersen 2009; Gow et al. 2010). Table 3, Panel A reports results using analyst recommendations and recommendation revision quintiles as the dependent variable, with quintiles coded from 1 to 5.¹¹

For recommendation levels, we find that analysts correctly incorporate the implications for future returns of only one of the *Accounting* variables: unexpected earnings (*SUE*). Analysts favorably recommend firms with high total accruals (*TACCR*), and do not consider capital expenditures (*CAPEX*), despite evidence that increases in those accounting measures are associated with lower future returns (Sloan 1996). Examining the *Valuation* measures, analysts correctly favor smaller firms (*LN MVE*) and those with a higher earnings-to-price ratio (*EP*). However, they also favor firms with a low book-to-market ratio (*BTM*) and high growth (*SG*, *LTG*), despite evidence that stock prices of such firms underperform the market. Examining the *Momentum* variables, analysts correctly favor firms with high earnings momentum (*FREV*) and stock price momentum (*MOM*).

The results for revisions in analysts' recommendations are reported on the right side of Table 3, Panel A. Examining the *Accounting* variables, we find that all three variables (*SUE*, *TACCR*, and *CAPEX*) are statistically significant, but with the unexpected sign. For the *Valuation* and *Growth* variables, the evidence is mixed: *EP* and *TURN* are statistically significant in the expected direction, but *BTM* is statistically significant in the unexpected direction. The coefficient on *LTG* is also significant in the unexpected direction. For the *Momentum* variables, we find that both *MOM* and *FREV* are statistically significant in the expected direction. Finally, we find that recommendation changes are negatively associated with past recommendation levels. This result makes intuitive sense because the highest (lowest) recommendations can only be revised down (up).

Considering the types of information used by analysts in both their recommendations and recommendation revisions, analysts' correctly favor stocks with positive price momentum (*MOM*), positive earnings momentum (*FREV*), and high earnings-to-price (*EP*). They incorrectly favor stocks with high forecasted growth (*LTG*), high accruals (*TACCR*), and low book-to-market value (*BTM*). Thus, financial analysts view higher past and future growth and higher accruals as positive features in recommending stocks, despite research that shows the opposite relation (Lakonishok et al. 1994; La Porta 1996; Sloan 1996). In addition, analysts also tend to issue more favorable recommendations for firms with low book-to-market ratios, even though prior research shows a positive association with subsequent returns (Fama and French 1992). This evidence indicates that

¹¹ Note that quintiles are of approximate equal size (after adjusting for ties and including all recommendation revisions of zero in the middle quintile). Due to the low frequency of strong sell and sell recommendations issued by analysts, the most unfavorable recommendation quintile contains some "hold" recommendations.

TABLE 3
Use of Predictive Information by Analysts and Short Sellers
Panel A: Explaining Recommendation Levels and Changes (Using Ordered Logistic Regression)

Variable	Predict	Recommendation Levels		Recommendation Changes	
		Coefficient	Chi-Square	Coefficient	Chi-Square
<i>Accounting</i>					
<i>SUE</i>	Pos	1.837	15.70***	-0.743	13.81***
<i>TACCR</i>	Neg	0.994	21.16***	0.424	4.52**
<i>CAPEX</i>	Neg	0.006	0.00	0.437	8.42***
<i>Valuation</i>					
<i>LnMVE</i>	Neg	-0.034	8.78***	0.008	1.34
<i>EP</i>	Pos	2.278	73.73***	0.564	20.31***
<i>BTM</i>	Pos	-0.714	119.05***	-0.170	31.79***
<i>TURN</i>	Neg	-0.032	0.17	-0.319	58.70***
<i>Growth</i>					
<i>SG</i>	Neg	0.781	213.28***	-0.015	0.15
<i>LTG</i>	Neg	0.041	202.87***	0.015	61.73***
<i>Momentum</i>					
<i>FREV</i>	Pos	8.254	115.68***	5.170	76.91***
<i>MOM</i>	Pos	0.473	33.60***	0.499	71.64***
<i>LAG_QRec</i>	Neg			-0.926	1052.85***
Pseudo R ²		0.138		0.098	

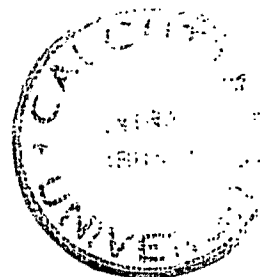
(continued on next page)

Panel B: Explaining Short Interest (Using Ordered Logistic Regression)

Variable	Predict	Short Interest	
		Coefficient	Chi-Square
<i>Accounting</i>			
<i>SUE</i>	Neg	-1.777	21.93***
<i>TACCR</i>	Pos	1.527	21.63***
<i>CAPEX</i>	Pos	1.974	32.66***
<i>Valuation</i>			
<i>LnMVE</i>	Pos	0.080	2.83*
<i>EP</i>	Neg	-0.698	18.04***
<i>BTM</i>	Neg	-0.245	14.46***
<i>TURN</i>	Pos	4.198	484.78***
<i>Growth</i>			
<i>SG</i>	Pos	0.410	45.46***
<i>LTG</i>	Pos	0.014	23.66***
<i>Momentum</i>			
<i>FREV</i>	Neg	-2.138	15.27***
<i>MOM</i>	Neg	-0.096	5.25**
Pseudo R ²		0.332	

*, **, *** Indicate statistical significance at the $\alpha = 0.10$, 0.05, and 0.01 levels, respectively, using a two-tailed test.
 n = 80,674 firm-quarters.

This table reports estimation results when analysts' recommendation and short interest quintile assignments are regressed (using ordered Logit) on 11 variables shown to be predictive of future returns. We do not report the intercepts for parsimony. See Table 1 for descriptions of each variable, and the Appendix for detailed explanations of how each variable is calculated. The "Predict" column reports the predicted relation between the explanatory variable and future returns as indicated in prior research. We report test statistics based on standard errors that are adjusted for two-way clustering of residuals by firm and calendar month.



sell-side analysts tend to favorably recommend “glamour stocks.” JKKL reached the same conclusion based on an analysis of an earlier time period.

Table 3, Panel B reports results from a model using short interest quintiles as the dependent variable. We find that all 11 variables are statistically significant with coefficient signs in the expected direction.¹² Additionally, we note that the explanatory power of this model (Panel B, pseudo $R^2 = 33.2$ percent) is more than double that for the model using recommendation levels (Panel A, pseudo R^2 of 13.8 percent) and more than triple that for the model using recommendation revisions (Panel A, pseudo R^2 of 9.8 percent).¹³ Thus, consistent with our univariate analysis, we find that short interest is explained better by the predictive information in *Accounting*, *Valuation*, and *Growth* variables than is analyst recommendation levels or changes. Our evidence is consistent with other studies that examine the association between short interest and indicators of future returns (Dechow et al. 2001; Cao et al. 2007; Seybert and Wang 2009). Our study extends this research by employing several predictive variables simultaneously in the same regression model, and by comparing results for short sellers to those for analysts.

Incremental Information About Future Returns

In this section, we investigate whether recommendations, recommendation changes, and short interest contain incremental information about future returns, beyond the information in the 11 predictive variables. Using the methodology in JKKL, we convert the continuous predictor variables into binary signals based on a median split. For all variables where the expected relation with future returns is positive (negative), the binary variable is coded 1 when its value is greater (less) than its median for a given quarter, and 0 otherwise. Thus, we expect a positive coefficient on all predictor variables.

Table 4 presents the results from our analysis. The model on the left includes both recommendation level quintiles (*QRec*) and short interest quintiles (*QSiratio*). Each of these quintiles is scaled to range from 0 to 1 to facilitate interpretation of coefficients. The coefficient on *QRec* indicates that analysts’ recommendations are incrementally informative about future abnormal returns, but the coefficient is negative. This result suggests that buy-and-hold investors would do better to trade against the consensus analyst recommendation. In contrast, the coefficient on *QSiratio* is significantly negative, indicating that short interest provides incremental information for predicting future returns even after controlling for the information contained in the 11 predictor variables. The negative sign indicates that, as would be expected, a higher (lower) level of short interest is associated with lower (higher) future abnormal returns. The coefficient magnitude for *QSiratio* of -0.029 can be interpreted as the six-month return earned on an investment portfolio that is formed optimally to exploit the information in short interest that is orthogonal to the information in analysts’ recommendations and the predictive variables. Thus, while short sellers use the information contained in the investment signals (as indicated by the significant associations reported in Table 3), the results in Table 4 show that short sellers also develop information to predict future returns that goes beyond what is contained in those variables. Examining the predictor variables, we find significant coefficients in the expected direction for *TACCR*, *SG*, and *FREV*.¹⁴

¹² Note that the explanatory variables have the opposite predicted sign in the short interest model (compared to the recommendation models).

¹³ Since the dependent variables differ across models, it is not possible to test for differences in explanatory power. However, given that we have standardized the dependent variables by ranking them into quintiles, their variation is similar. Specifically, the standard deviations of the quintile ranking of analyst levels, analyst changes, and short interest are 1.41, 1.38, and 1.41, respectively. Thus, we believe a comparison of pseudo R^2 s is informative.

¹⁴ The coefficient for *TURN* is marginally significant with the wrong sign. This appears to be driven by its high correlation with short interest (greater than 50 percent). When *QSiratio* is excluded from the model, the coefficient for *TURN* is not

TABLE 4
**Incremental Information about Future Returns Provided by Recommendations,
 Recommendation Revisions, and Short Interest**

Variable	Recommendations and Short Interest		Recommendation Changes and Short Interest	
	Coefficient	t-stat	Coefficient	t-stat
Intercept	0.009	0.70	-0.003	-0.31
<i>QRec</i>	-0.019	-2.55**		
<i>QChgRec</i>			0.003	0.59
<i>QSIratio</i>	-0.029	-3.91***	-0.027	-3.77***
<i>DSUE</i>	0.004	0.79	0.003	0.63
<i>DTACCR</i>	0.025	7.10***	0.025	7.23***
<i>DCAPEX</i>	0.004	0.75	0.004	0.72
<i>DLnMVE</i>	0.003	0.72	0.003	0.68
<i>DEP</i>	-0.003	-0.45	-0.004	-0.54
<i>DBTM</i>	0.005	0.93	0.007	1.19
<i>DTURN</i>	-0.012	-1.85*	-0.012	-1.80*
<i>DSG</i>	0.015	2.43**	0.016	2.66***
<i>DLTG</i>	-0.009	-1.08	-0.007	-0.85
<i>DFREV</i>	0.011	2.05**	0.009	1.75*
<i>DMOM</i>	0.009	1.13	0.007	0.94
Adj-R ²	0.003		0.003	

*, **, *** Indicate statistical significance at the $\alpha = 0.10, 0.05$, and 0.01 levels, respectively, using a two-tailed test.
 n = 80,674 firm-quarters.

This table reports estimation results when future six-month abnormal returns (calculated as in Daniel et al. [1997]) are regressed on analysts' recommendations and short interest data along with 11 variables that prior research shows to be predictive of future returns. *QRec* is the quintile assignment based on recommendation levels. *QChgRec* is the quintile assignment based on recommendation revisions. *QSIratio* is the quintile assignment based on short interest. *QRec*, *QChgRec*, and *QSIratio* are scaled to range between 0 and 1 (0.00, 0.25, 0.50, 0.75, 1.00) to facilitate the interpretation of the coefficients. Using the methodology in Jegadeesh et al. (2004), we convert the continuous predictor variables into binary signals based on a median split. For all variables where the expected relation with future returns is positive (negative), the binary variable is coded 1 when its value is greater (less) than its median for a given quarter, and 0 otherwise. See Table 1 for descriptions of each variable, and the Appendix for detailed explanations of how each variable is calculated. We report test statistics based on standard errors that are adjusted for two-way clustering of residuals by firm and calendar month.

The model on the right side of Table 4 provides a similar analysis for recommendation revision quintiles (*QChgRec*) and short interest (*QSIratio*) quintiles. The coefficient on *QChgRec* is positive as expected, but is not significantly different from zero, suggesting that recommendation revisions do not provide information about future returns that is incremental to short interest and to the other publicly available investment signals. The coefficient on *QSIratio* is again significantly negative. The coefficients and significance levels for the predictor variables are similar to those reported in the recommendation level regressions.¹⁵

statistically significant (t-stat = -0.52). In contrast, *QSIratio* remains highly significant when *TURN* is excluded from the model.

¹⁵ We also estimated a regression equation that includes *QRec*, *QChgRec*, and *QSIratio* together with the other 11 predictive variables. Results from this regression are qualitatively equivalent to what is reported in Table 4, except that

We draw the following general conclusions from the results presented in Table 3 and Table 4. First, analysts' recommendations do a poor job of incorporating information about future returns provided by the predictor variables, and buy-and-hold investors would actually do better by trading against analyst recommendations. Second, analyst recommendation revisions also do a poor job of using predictive information, and they do not contribute information beyond what is contained in the predictor variables and short interest. Third, short sellers correctly incorporate publicly available information that is predictive of future returns and, furthermore, short sellers are able to generate information that is orthogonal to the set of predictive variables we use in our analysis. In the next section, we explore the success of trading strategies that are designed to exploit the above results.

IV. INVESTMENT PERFORMANCE BASED ON ANALYST RECOMMENDATIONS AND SHORT INTEREST

The results reported in Section III suggest that investors might improve their returns by using short interest as a supplementary investment signal. Indeed, many of the associations between analyst recommendations and variables that are predictive of future returns suggest that analysts might actually impede the price discovery process. Pownall and Simko (2005) provide evidence that short positions can help to fill an information intermediary gap for companies with low analyst following. Our results suggest that short interest can be used as an information signal for investors, regardless of the analyst following. The benefit of short interest is likely to be greatest when this signal contradicts the signal from analysts' recommendations. Specifically, the results in Table 4 suggest that investors would profit from a buy-and-hold strategy that (1) trades against analyst recommendation levels and (2) trades with short interest. This conjecture is based on the signs of the statistically significant coefficients for analyst recommendations and short interest in Table 4.¹⁶

The trading strategies we consider are implementable and follow the portfolio construction methodology outlined in Jegadeesh and Titman (1993).¹⁷ Under this methodology, the strategies hold a series of sub-portfolios that formed in the current month and in each of the previous five months (six-month holding period).¹⁸ Thus, we simulate a portfolio where a 1/6 fraction of the stocks are reassigned to portfolios each month. We rebalance the portfolios monthly to maintain equal weights on each security and calculate the mean abnormal return for each portfolio. This results in a time-series of monthly portfolio returns that is free of overlapping return accumulation periods. As discussed in further detail below, we also calculate hedge portfolio returns that go long and short in particular portfolios.

Since the consensus analyst recommendation and the short interest ratio change each month, each portfolio is based on information from analysts and/or short sellers from the current month and from each of the past five months. In our initial tests, we examine the signals from analysts or short sellers separately. We then combine recommendations and short interest to develop a trading strategy that exploits the information contained in both signals. When combining the signals, we intersect the analyst recommendation and short interest quintiles to produce 25 portfolios (5×5).

the coefficient on *QChgRec* becomes marginally significantly positive.

¹⁶ In this section, we do not report trading strategies that use recommendation revisions because results from these tests are generally consistent with our earlier analyses that show they contain no incremental information about future returns beyond short interest.

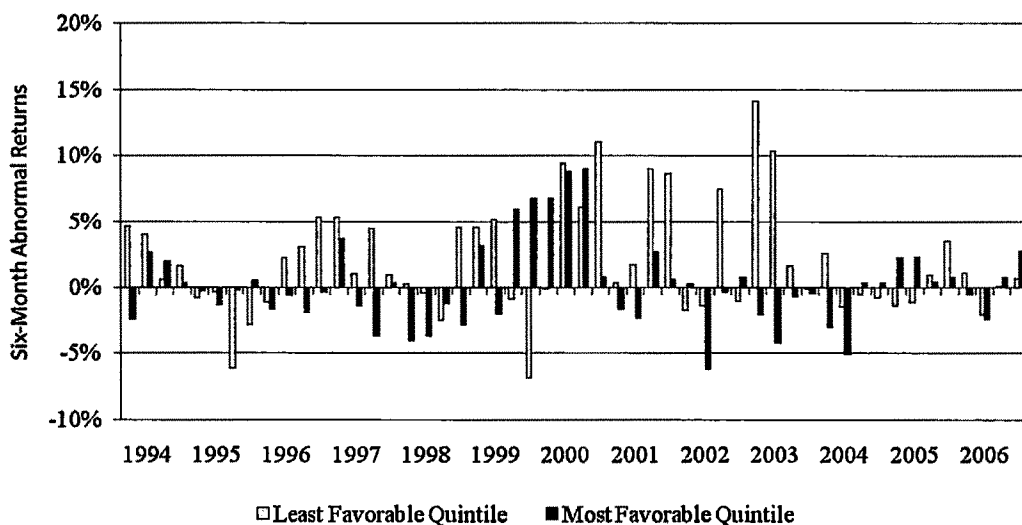
¹⁷ The Jegadeesh and Titman (1993) methodology avoids statistical problems associated with serial correlation induced by overlapping return accumulation periods. We thank an anonymous reviewer for suggesting this approach.

¹⁸ In a later section, we discuss the sensitivity of the results to a shorter return window of one month and the use of a four-factor model, which controls for the Fama-French risk factors and momentum.

We form our portfolios using all firm-month observations with available data for analyst recommendations, short interest levels, and stock returns. Thus, we drop the requirement that data are available for the 11 predictor variables, and we include every month (not just the last month of each quarter, as in our first set of analyses). These changes increase the sample size to 564,101 firm-month observations and, as a result, increase the extent to which we can generalize results.¹⁹ We report results for the full time period, 1994 to 2006, and for three sub-periods because documenting the stability of abnormal returns to a particular trading strategy over time is critical to the evaluation of the strategy. The first sub-period consists of the five-year period from 1994 to 1998, which overlaps the 1985 to 1998 time period investigated by JKKL, and corresponds to a strong bull market. The second sub-period consists of the subsequent five years, from 1999 to 2003, which includes the precipitous decline in the NASDAQ index, the adoption of Regulation Fair Disclosure, and the Global Settlement agreement reached between the SEC, NASD, NYSE, and ten of the largest investment firms. The final sub-period consists of the most recent years in our sample, 2004 to 2006.

In Figure 1, we display six-month abnormal quarterly returns over the full test period for

FIGURE 1
Comparison of Six-Month Abnormal Returns to the Most and Least Favorable Recommendations



Consensus recommendations as of the last month of calendar quarter t are sorted into quintiles, with the highest quintile designated the most favorable portfolio and the lowest quintile the least favorable portfolio. Portfolios are reformed each calendar quarter and we report abnormal buy-and-hold returns for the six months beginning the first day of quarter $t+1$.

¹⁹ In a robustness test (untabulated), we find that inferences from our trading strategies are unchanged when we restrict the sample to include firms *with* available data for the 11 predictor variables.

stocks in the least favorable recommendation quintile and for those in the most favorable recommendation quintile. Our first observation is that the variability of the absolute value of the abnormal returns differs considerably among the three sub-periods. The variability is much higher in the 1999–2003 sub-period than in the preceding 1994–1998 sub-period; return variability then drops precipitously and remains relatively low throughout the 2004–2006 sub-period. Our second observation is that analyst recommendations are not very reliable as a predictor of future returns. In fact, stocks in the least favorable recommendation quintile earn greater returns than those in the most favorable quintile in 35 of the 72 quarters. So, flipping a coin appears to be just as predictive.

Returns for Portfolios Based on Analyst Recommendations or Short Interest

Table 5, Panel A reports average monthly abnormal returns based on analyst recommendations. The portfolios cover the 52 calendar quarters from 1994 to 2006. If analyst recommendations provide value to investors, then returns should be higher for more highly recommended stocks. That is, returns for stocks in the quintile with the highest average recommendation should exceed returns for lower recommendation quintiles. Consistent with the results previously presented in Table 4 (which indicate that analysts improperly incorporate value-relevant information into their forecasts), we find the opposite relation. We find that the mean abnormal returns generally *decrease* as the recommendation level increases. When we calculate returns from investing long in firms classified in the most favorable recommendation quintile and taking an offsetting short position in firms in the least favorable recommendation quintile, we obtain a *modestly negative* return of 26 monthly basis points ($t = -2.03$; $p < 0.01$).

Table 6, Panel A summarizes returns by recommendation quintile for each of the three sub-periods and provides statistical tests. The investment results vary considerably among the three sub-periods. However, the most favorable recommendation quintile does not perform better than fourth among the quintiles (see ranks in the right-most column). Two statistically significant portfolios fall in the 1999–2003 period, when returns are positive for the lowest recommendation quintiles. We also find that the hedge portfolio trading strategy (buying firms in the highest recommendation quintile and selling short those in the lowest recommendation quintile) yields a statistically significant return only in the 1999–2003 sub-period, and that return is a *negative* 48 monthly basis points. Overall, following the consensus analyst recommendation does not generate positive abnormal returns for investors and, in some periods, can actually generate significantly negative returns. While perhaps counterintuitive, this result is consistent with Barniv et al. (2009), who find that analysts' recommendations relate negatively to residual income valuation models, and with Seybert and Wang (2009), who find that firms with more optimistic recommendations earn lower future returns in periods of high investor sentiment.²⁰

Table 5, Panel B reports average monthly abnormal returns from portfolios formed using short interest levels. We find that the average abnormal return across the quintiles declines monotonically from lowest to highest short interest. A trading strategy that invests long (short) in firms that are in the lowest (highest) short interest quintile provides an average monthly abnormal return of 56 basis points, which is both economically and statistically significant ($t = 2.46$; $p < 0.01$). Interestingly, 29 basis points of that return are earned because short sellers avoid sizable positions in stocks that yield positive future returns. Thus, not only are short sellers able to identify stocks that are likely to fall in price, but they successfully avoid stocks that are likely to realize price

²⁰ Relative to JKKL, we find a more negative relation between analyst recommendations (both levels and changes) and future returns. Our sub-period analysis suggests that an important factor driving this result is the different time periods examined across the two studies.

TABLE 5
Abnormal Returns to Portfolios Based on Analyst Recommendations and Short Interest
1994–2006

Panel A: Average Monthly Abnormal Returns Based on Recommendation Levels

Quintile Portfolios	Performance 1994 to 2006		Quintile Rank by Return
	Mean Return	t-stat	
Most Favorable Recommendations	-0.10%	-1.37	5 (Worst)
	0.01%	0.13	4
	0.11%	2.32**	3
Least Favorable Recommendations	0.19%	3.64***	1 (Best)
	0.16%	2.15**	2
Most Favorable – Least Favorable	-0.26%	-2.03**	

Panel B: Average Monthly Abnormal Returns Based on Short Interest

Quintile Portfolios	Performance 1994 to 2006		Quintile Rank by Return
	Mean Return	t-stat	
Lowest Short Interest	0.29%	3.19***	1
	0.20%	3.64***	2
	0.12%	2.70***	3
Highest Short Interest	0.04%	0.63	4
	-0.27%	-1.79*	5
Lowest SI – Highest SI	0.56%	2.46**	

*, **, *** Indicate statistical significance at the $\alpha = 0.10$, 0.05, and 0.01 levels, respectively, using a two-tailed test.
 n = 564,101 firm-months.

We report mean monthly abnormal returns adjusted for firm size, book-to-market ratio, and past returns (Daniel et al. 1997) for portfolios based on analyst recommendations or short interest. Using the methodology in Jegadeesh and Titman (1993), we construct portfolios monthly based on the quintile rank of analyst recommendations or short interest selected in the current month and in each of the past five months (six-month holding period).

TABLE 6
Abnormal Returns to Portfolios Based on Analyst Recommendations or Short Interest for Sub-Periods

Panel A: Average Monthly Abnormal Returns Based on Recommendation Levels

Quintile Portfolios	Performance 1994 to 1998			Performance 1999 to 2003			Performance 2004 to 2006		
	Mean Return	t-stat	Rank by Return	Mean Return	t-stat	Rank by Return	Mean Return	t-stat	Rank by Return
Most Favorable Recommendations	-0.16%	-1.53	5	-0.17%	-1.22	5	0.02%	0.26	4
	-0.07%	-0.58	4	0.03%	0.23	4	0.05%	0.75	3
	0.11%	1.68*	2	0.15%	1.53	3	0.06%	1.00	1
	0.10%	1.52	3	0.40%	3.93***	1	0.06%	0.80	2
Least Favorable Recommendations	0.13%	1.16	1	0.31%	2.11**	2	-0.06%	-0.63	5
Most Favorable - Least Favorable	-0.29%	-1.58		-0.48%	-1.93*		0.09%	0.55	

Panel B: Average Monthly Abnormal Returns Based on Short Interest

Quintile Portfolios	Performance 1994 to 1998			Performance 1999 to 2003			Performance 2004 to 2006		
	Mean Return	t-stat	Rank by Return	Mean Return	t-stat	Rank by Return	Mean Return	t-stat	Rank by Return
Lowest Short Interest	0.22%	1.72*	1	0.30%	1.69*	1	0.30%	2.74***	1
	0.17%	2.09**	2	0.27%	2.52**	2	0.11%	1.61	2
	0.05%	0.93	3	0.19%	2.17**	3	0.05%	0.79	3
	0.01%	0.10	4	0.18%	1.31	4	-0.08%	-0.97	4
Highest Short Interest	-0.30%	-1.43	5	-0.24%	-0.80	5	-0.24%	-1.59	5
Lowest SI - Highest SI	0.52%	1.63		0.54%	1.20		0.54%	2.43***	

*, **, *** Indicate statistical significance at the $\alpha = 0.10$, 0.05, and 0.01 levels, respectively, using a two-tailed test.

n = 564,101 firm-months.

We report mean monthly abnormal returns adjusted for firm size, book-to-market ratio, and past returns (Daniel et al. 1997) for portfolios based on analyst recommendations or short interest. Using the methodology in Jegadeesh and Titman (1993), we construct portfolios monthly based on the quintile rank of analyst recommendations or short interest selected in the current month and in each of the past five months (six-month holding period).

increases. This finding is consistent with concurrent research by Boehmer et al. (2010), who also find that the positive signals from low short interest can be equal to or greater in absolute magnitude than the negative signal from high short interest.

Table 6, Panel B presents returns from trading on short interest levels for each of the three sub-periods. A monotonic pattern of decreasing returns occurs within each of the three sub-periods as one reads down the table from lowest to highest short interest, with several of the individual quintiles statistically significant. Over the 2004–2006 sub-period, the zero-investment hedge strategy produces a statistically significant abnormal return of 54 monthly basis points. The monthly hedge returns are similar for the other sub-periods but are not statistically significant. Thus, while larger abnormal returns can be earned by trading on the level of short interest rather than on analysts' recommendations, the hedge strategy results are not of sufficient magnitude to be statistically significant in each sub-period. Interestingly, the returns to buying stocks in the lowest short interest quintile are more likely to be statistically significant than are the returns to selling short the stocks in the high short interest quintile. Overall, short interest provides a more reliable signal about which stocks to *buy* than about which stocks should be sold short.

Returns for Portfolios Combining Analyst Recommendations and Short Interest Levels

The evidence presented to this point indicates that none of the signals in isolation produce consistent results across each sub-period. In this section, we investigate whether combining the signals from analysts and short sellers can improve upon this investment performance. We consider trading strategies that are based on concurring and conflicting signals from analysts and short sellers. To conduct our analysis, we independently sort analysts' recommendations and short interest into quintiles and merge these quintile rankings to form 25 different portfolios. Since the sorts are independent, the average number of stocks across portfolios is not equal. Nevertheless, as reported in Table 7, the sample is broadly distributed across the 25 different portfolios, ranging from an average number of firms per portfolio of 492 to 950. This suggests that analysts and short sellers are just as likely to disagree as they are to agree on the future prospects of any one firm.

Table 7 presents abnormal returns for portfolios formed using both analyst recommendations and short interest. Reading down each recommendation column in Table 7, Panel A shows that abnormal returns tend to decline as the level of short interest increases. Notably, the decline is monotonic for the three quintiles with the most favorable recommendations. In addition, reading across each short interest row from left to right shows a general tendency for abnormal returns to decrease as the recommendation becomes more favorable. The combined effect is that the lowest returns occur in the lower right quadrant and the highest in the upper left quadrant.

Table 7, Panel B reports returns for two zero-investment trading strategies. The first strategy is to trade when short sellers and analysts strongly concur about a company's prospects. The specific trade is to buy firms in the portfolio with the most favorable recommendations and lowest short interest, and to sell short firms with the worst recommendations and highest short interest. The upper half of Panel B reports the returns and t-statistics from this strategy. The sample period return is 36 monthly basis points, which is not statistically significant ($t = 1.25$; $p = 0.21$). Examining sub-periods, a positive abnormal return occurs in each sub-period; however, the return is only statistically significant during the 2004 to 2006 time period (at the 10 percent level). We also note that the returns in each sub-period are less than the returns available to investors by trading only on the level of short interest (see Table 6, Panel B).

The second strategy is to trade when short sellers and analysts strongly conflict about a company's prospects. The strategy is to follow the short sellers by buying firms in the portfolio with the least favorable recommendations but the lowest short interest level, while selling short firms with the best recommendations but the highest short interest. The combined return reported in Panel B is 111 monthly basis points ($40 + 71$ from the corner cells in Panel A), which is both

TABLE 7
Abnormal Returns to Portfolios Formed Using Recommendations and Short Interest
1994–2006

Panel A: Average Monthly Returns (Portfolio Sample Size in Parentheses) for Cross-Tabulation of Recommendation Levels and Short Interest

Short Interest Quintiles	Analyst Recommendation Quintiles, 1994–2006				
	Least Favorable				Most Favorable
Lowest Short Interest	0.40%*** (950)	0.19% (492)	0.31%*** (757)	0.29%*** (601)	0.24%*** (932)
	0.19%* (799)	0.27%*** (691)	0.30%*** (820)	0.11% (691)	0.14%* (788)
	0.20%* (723)	0.20%*** (829)	0.23%*** (844)	0.03% (765)	–0.07% (696)
	0.16% (700)	0.26%*** (874)	0.00% (835)	–0.03% (821)	–0.18% (665)
Highest Short Interest	–0.12% (667)	0.06% (795)	–0.19% (781)	–0.32%* (874)	–0.71%*** (728)

(continued on next page)

Panel B: Average Monthly Hedge Portfolio Returns

When Short Interest and Analyst Recommendations Strongly Concur

	1994-1998		1999-2003		2004-2006		Sample Period 1994-2006	
	Mean	t-stat	Mean	t-stat	Mean	t-stat	Mean	t-stat
Firms with lowest short interest and most favorable analyst recommendation = Buy	0.01%	0.04	0.25%	1.04	0.39%	2.51**	0.24%	2.12**
Firms with highest short interest and least favorable analyst recommendation = Sell	-0.31%	-1.14	0.12%	0.24	-0.11%	-0.49	-0.12%	-0.55
Buy-Sell	0.30%	0.88	0.13%	0.22	0.50%	1.74*	0.36%	1.25

When Short Interest Strongly Conflicts with Analyst Recommendations

Firms with lowest short interest but least favorable analyst recommendation = Buy	1.95*	0.39%	1.45	0.25%	1.45	0.40%	2.81***
Firms with highest short interest but most favorable analyst recommendation = Sell	-2.68***	-0.91%	-2.95**	-0.46%	-2.11**	-0.71%	-4.27***
Buy - Sell	1.10%	2.58**	1.30%	2.62**	0.71%	2.20**	4.09***

*, **, *** Indicate statistical significance at the $\alpha = 0.10$, 0.05, and 0.01 levels, respectively, using a two-tailed test.

n = 564,101 firm-months.

We report mean monthly abnormal returns adjusted for firm size, book-to-market ratio, and past returns (Daniel et al. 1997) for portfolios based on analyst recommendations and short interest. In parentheses, we report the average number of firms per month in each portfolio. Using the methodology in Jegadeesh and Titman (1993), we construct portfolios monthly based on the quintile rank of analyst recommendations and short interest selected in the current month and in each of the past five months (six-month holding period).

sizable and statistically significant ($t = 4.09$; $p < 0.01$). Positive returns accrue on both the long and short side, and the combined return is about double the return of 54 basis points available from trading only on the level of short interest (see Table 5, Panel B). This trading strategy also produces statistically significant returns of 110 basis points in 1994–1998, 130 basis points in 1999–2003, and 71 basis points in 2004–2006, providing evidence of the stability of this strategy over time. Note that the highest return occurs in the 1999–2003 time period when analyst recommendations are most misleading and produce a negative return of -48 monthly basis points (Table 6, Panel A).

In sum, we find that combining investment signals from analysts and short sellers yields incrementally greater future returns than trading strategies that use only one of the signals. Consistent with the results presented in Table 4, the most profitable investment strategy is when an investor trades in firms about which analysts and short sellers strongly disagree, and the investor takes the buy or sell position indicated by the short interest signal.

Robustness Tests

In this section, we report the results of several robustness tests (all untabulated). We begin by examining the sensitivity of the results to using the I/B/E/S provided consensus recommendation, rather than our self-constructed consensus. We find that our sample size increases slightly (to 86,592 firm-quarter observations) with all results qualitatively the same as those reported.

Next, we use an alternative short interest variable. Recall that for our main analyses, we use the short interest ratio (open short interest divided by shares outstanding). As an alternative deflator, we scale open short interest by the previous month's trading volume and label this variable *SIVOL*. We find that *SIVOL* is highly correlated with the short interest ratio ($\rho = 0.74$). When we regress the quintile assignment of *SIVOL* on the 11 predictive variables, we again find that all 11 variables are statistically significant in the expected direction, which is consistent with the results using the short interest ratio.

Our remaining robustness tests reexamine the profitability of the trading strategies that use information from both analysts and short sellers. First, we examine whether our results are robust to monthly rebalancing of the portfolio and using a holding period of one month. Consistent with our main results, we find that the most profitable strategy is to follow the short sellers when the signals from analysts and short sellers strongly conflict (100 monthly basis points; $t = 3.39$; $p < 0.01$). The alternative strategy, which trades when analyst and short seller signals strongly concur, remains less profitable although it improves from 36 monthly basis points ($t = 1.25$; $p = 0.21$) to 75 basis points ($t = 2.14$; $p < 0.01$).

Second, we examine whether the returns to our hedge portfolios become more profitable when we use a finer partition to form portfolios. Each month, we sort firms into deciles (instead of quintiles) based on the consensus analyst recommendation and/or the level of short interest. We intersect the recommendation and short interest deciles to produce 100 portfolios and the strategies take positions in the four most extreme portfolios using a six-month holding period. We find that the more refined stock selection yields greater hedge portfolio returns (140 monthly basis points; $t = 3.35$; $p < 0.01$).

Finally, we estimate the returns to our trading strategies using calendar-time portfolios, as an alternative method for estimating abnormal returns. Specifically, we assess the profitability of the strategies by estimating alphas from a regression of each portfolio's time-series of excess returns on the Fama-French risk factors and momentum (Fama and French 1993; Carhart 1997). We find that all hedge portfolio returns maintain statistical significance, which is consistent with the main results that calculate abnormal returns using characteristic portfolio matching (Daniel et al. 1997). We again find that the most profitable strategy is to follow the short sellers when their positions conflict with analyst recommendations.

V. CONCLUSION

We contribute new findings on the characteristics of stocks favored by short sellers, about how those characteristics differ from those used by analysts in developing buy-hold-sell recommendations, and on the value relevance of short interest data to investors. By so doing, we expand upon the results of Dechow et al. (2001) and other studies that examine information used by short sellers.²¹ First, we find that analysts and short sellers use publicly available information differently. Analysts over-recommend stocks with high growth, high accruals, and low book-to-market ratios, even though prior research shows these characteristics are negatively related to future returns. In contrast, short sellers incorporate into their investment decisions the future return implications of all 11 accounting and market variables considered in this study. Second, we find that short interest provides information about future returns beyond that provided in the 11 items of information that prior research shows to be predictive of future returns. Analysts' recommendations also provide incremental information, but a negative coefficient suggests trading against the analysts. Third, based on these results, we show that a highly profitable trading strategy is one where investors trade with the short sellers when the short interest signal strongly conflicts with the consensus analyst recommendation. In fact, the value of short interest in choosing stocks to buy or sell is greater when conditioned on a conflicting consensus recommendation than when used by itself to trade stocks.

Our study contributes to the stream of academic literature documenting market inefficiencies. The debate over market efficiency has shifted from simple yes-or-no questions to issues such as the types of information incorporated into prices with a delay, the speed of price adjustment, and factors that facilitate or impede price discovery. We consider fundamental and other predictive information that prior research has shown to be incorporated into prices with a delay. We show that analysts' recommendations can impede price discovery of this information, but short selling facilitates price discovery. A frequent criticism of research documenting a delayed reaction to information is that a relation between information and future returns found to exist in the past may not recur in the future. That is, while some stocks are always misvalued, the characteristics of those stocks change over time, reducing the effectiveness of any system that places fixed weights on information. This criticism is less likely to be true for our approach, which identifies misvalued stocks based on a (strong) difference of opinion between analysts and short sellers. The characteristics of misvalued stocks can change over time, as long as short sellers and analysts disagree about valuation and the current stock price under-weights the views of short sellers. The extent of disagreement is also likely to be greatest in periods of high return volatility, and this is the type of market environment in which investors want stock-picking guidance. Consistent with this conjecture, we find the highest returns during the 1999–2003 sub-period.

Our study is timely in light of recent actions in the U.S. and other countries to further regulate short selling. For example, SEC Release No. 34-58591 (SEC 2008) requires institutional managers with at least \$100 million under management to report detailed information about daily short sales in new Form SH. The rationale in Release No. 34-58591 is that "sudden and unexplained declines in the prices of securities ... can give rise to questions about the underlying financial condition of an issuer, which in turn can create a crisis of confidence without a fundamental underlying basis."²² While our evidence does not include the specific time period referred to in this quote, any new regulations are likely to extend to more normal periods of market activity. An important implication of our study is that regulations that restrict or increase the cost of short selling run the

²¹ Two concurrent working papers consider aspects of short seller information use (Cao et al. 2007; Seybert and Wang 2009).

²² In SEC Release No. 58724 (October 2, 2008), the SEC states the daily short selling information reported on Form SH will not be publicly available, in part, because it could give rise to imitative short selling.

risk of limiting a potentially important source of information for investors about future equity values. In this regard, the SEC has recently taken actions to increase the public availability of short interest positions. On March 6, 2007, the SEC approved rule changes that increase the frequency of short interest reporting from monthly to twice a month, effective September 2007. More timely reporting of short interest data to the public should further increase the role of short sellers as an information intermediary.

APPENDIX

QUANTITATIVE INVESTMENT SIGNALS

The last month of each calendar quarter is labeled quarter t . On this date, we measure our stock recommendation and short interest variables. Relative to this date, we label as quarter q the most recent fiscal quarter for which an earnings announcement is made at least two months prior to the end of quarter t and no more than four quarters prior to the end of quarter t .

Variable	Description	Calculation Details	Normative correlation with subsequent returns
<i>SUE</i>	Unexpected earnings	Seasonally adjusted earnings scaled by price for fiscal quarter q , as calculated by: $\frac{EPS_q - EPS_{q-4}}{Price_q}$ where EPS = earnings per share before extraordinary items (DATA#19) divided by the split adjustment factor (DATA#17) [Compustat]; and $Price$ = stock price (DATA#14) divided by the split adjustment factor (DATA#17) [Compustat].	Positive (Bernard and Thomas 1989)
<i>TACCR</i>	Total accruals	Earnings before extraordinary items and discontinued operations (DATA#76) minus cash flow from operations (DATA#108 - DATA#78), scaled by average assets (DATA#44) as measured at the end of fiscal quarter q [Compustat]. Since Compustat reports cumulative (i.e., year-to-date) data for cash flow items, adjustments were made to arrive at total accruals for fiscal quarter q (see Collins and Hribar 2000).	Negative (Sloan 1996)
<i>CAPEX</i>	Capital expenditures	Rolling sum of the preceding four quarters of capital expenditures ending at fiscal quarter q divided by average total assets as calculated by: $\frac{\sum_{i=0}^3 Capex_{q-i}}{(TA_q - TA_{q-4}) / 2},$ where $Capex$ = capital expenditures (DATA#90); and TA = total Assets (DATA#44).	Negative (Beneish et al. 2001)
<i>MVE</i>	Market capitalization	Natural log of the market value of equity at the end of fiscal quarter q , as calculated by $DATA\#14 \times DATA\#61$ [Compustat].	Negative (Fama and French 1992)
<i>EP</i>	Earnings-to-price ratio	Ratio of the rolling sum of earnings over the preceding four quarters divided by price at the end of fiscal quarter q , as calculated by: $\sum_{i=0}^3 \frac{EPS_{q-i}}{Price_q},$ where EPS = earnings per share before extraordinary items (DATA#19) divided by the split adjustment factor (DATA#17) [Compustat]; and $Price$ = stock price (DATA#14) divided by the split adjustment factor (DATA#17) [Compustat].	Positive (Fama and French 1992)
<i>BTM</i>	Book-to-market ratio	Ratio of the book value of equity to the market value of equity at the end of fiscal quarter q , as calculated by $DATA\#59 / (DATA\#14 \times DATA\#61)$ [Compustat].	Positive (Fama and French 1992)

(continued on next page)

Variable	Description	Calculation Details	Normative correlation with subsequent returns
<i>TURN</i>	Stock turnover	Average daily volume turnover ratio measured as the exchange-specific, percentile rank of: $\sum_{i=1}^n \frac{\text{Daily Vol.} / \text{ShROUT}}{n},$ where <i>Daily Vol.</i> = daily stock volume [CRSP]; <i>ShROUT</i> = shares outstanding [CRSP]; and <i>n</i> = the number of trading days for the six-month period ending on the last trading day of calendar quarter <i>t</i>.	Negative (Lee and Swaminathan 2000)
<i>SG</i>	Sales growth	Rolling sum of the preceding four quarters of sales ending at fiscal quarter <i>q</i> divided by the rolling sum of the preceding four quarters of sales ending on quarter <i>q</i> - 1, as calculated by: $\frac{\sum_{i=0}^3 \text{Sales}_{q-i}}{\sum_{i=0}^3 \text{Sales}_{q-1-i}},$ where <i>Sales</i> = DATA#2 [Compustat].	Negative (Lakonishok et al. 1994)
<i>LTG</i>	Long-term growth forecast	Mean, consensus long-term earnings growth forecast at the end of calendar quarter <i>t</i> [I/B/E/S].	Negative (Lakonishok et al. 1994; La Porta 1996)
<i>FREV</i>	Forecast revision	Rolling sum of the preceding six-month earnings forecast revisions to price ratios, as calculated by: $\sum_{i=0}^5 \frac{\text{FEPS}_{m-i} - \text{FEPS}_{m-i-1}}{\text{Price}_{m-i-1}},$ where <i>FEPS</i> = mean, consensus analyst forecast for one-year-ahead (FY1) earnings-per-share [I/B/E/S]; <i>m</i> = the last month of calendar quarter <i>t</i>; and <i>Price</i> = stock price just prior to the consensus measurement date [I/B/E/S].	Positive (Bernard and Thomas 1989; Chan et al. 1996)
<i>MOM</i>	Stock momentum	Buy-and-hold raw stock return for six-month period ending one month prior to the end of quarter <i>t</i> [CRSP].	Positive (Jegadeesh and Titman 1993)

REFERENCES

- Asquith, P., and L. Meulbroek. 1995. An empirical investigation of short interest. Working paper. Massachusetts Institute of Technology.
- , M. B. Mikhail, and A. S. Au. 2005a. Information content of equity analyst reports. *Journal of Financial Economics* 75 (2): 245–282.
- , P. A. Pathak, and J. R. Ritter. 2005b. Short interest, institutional ownership, and stock returns. *Journal of Financial Economics* 78 (2): 243–276.
- Barber, B. M., R. Lehavy, M. McNichols, and B. Trueman. 2001. Can investors profit from the prophets? Security analyst recommendations and stock returns. *The Journal of Finance* 56: 531–563.
- , ———, and B. Trueman. 2007. Comparing the stock recommendation performance of investment banks and independent research firms. *Journal of Financial Economics* 85 (2): 490–517.
- , ———, and ———. 2010. Ratings changes, ratings levels, and the predictive value of analysts' recommendations. *Financial Management* 39 (2): 533–553.
- Barniv, R., O.-K. Hope, M. Myring, and W. B. Thomas. 2009. Do analysts practice what they preach and should investors listen? Effects of recent regulations. *The Accounting Review* 84 (4): 1015–1039.
- Beneish, M. D., C. M. C. Lee, and R. L. Tarpley. 2001. Contextual fundamental analysis through the prediction of extreme returns. *Review of Accounting Studies* 6 (2–3): 165–189.
- Bernard, V. L., and J. K. Thomas. 1989. Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research* 27 (Supplement): 1–36.
- Boehmer, E., Z. R. Huszar, and B. D. Jordan. 2010. The good news in short interest. *Journal of Financial Economics* 96 (1): 80–97.
- , C. M. Jones, and X. Zhang. 2008. Which shorts are informed? *The Journal of Finance* 63 (2): 491–527.
- . 2009. Shackling short sellers: The 2008 shorting ban. Working paper. University of Oregon.
- Bradshaw, M. T. 2004. How do analysts use their earnings forecasts in generating stock recommendations? *The Accounting Review* 79 (1): 25–50.
- , S. A. Richardson, and R. G. Sloan. 2001. Do analysts and auditors use information in accruals? *Journal of Accounting Research* 39 (1): 45–74.
- Cao, B., D. Dhaliwal, A. C. Kolasinski, and A. V. Reed. 2007. Bears and numbers: Investigating how short sellers exploit and affect earnings-based pricing anomalies. Working paper. University of Washington.
- Carhart, M. M. 1997. On persistence in mutual fund performance. *The Journal of Finance* 52 (1): 57–82.
- Chan, L. K. C., N. Jegadeesh, and J. Lakonishok. 1996. Momentum strategies. *The Journal of Finance* 51 (5): 1681–1713.
- Collins, D. W., and P. Hribar. 2002. Errors in estimating accruals: Implications for empirical research. *Journal of Accounting Research* 40 (1): 105–134.
- Daniel, K., M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of Finance* 52 (3): 1035–1058.
- Dechow, P. M., A. P. Hutton, L. Meulbroek, and R. G. Sloan. 2001. Short-sellers, fundamental analysis, and stock returns. *Journal of Financial Economics* 61 (1): 77–106.
- Demirakos, E., N. Strong, and M. Walker. 2004. What valuation models do analysts use? *Accounting Horizons* 18 (4): 221–240.
- Desai, H., K. Ramesh, S. R. Thiagarajan, and B. V. Balachandran. 2002. An investigation of the informational role of short interest in the NASDAQ market. *The Journal of Finance* 57 (5): 2263–2287.
- Diamond, D. W., and R. E. Verrecchia. 1987. Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics* 18 (2): 277–311.
- Fama, E. F., and K. R. French. 1992. The cross-section of expected stock returns. *The Journal of Finance* 47: 427–465.
- . 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33 (1): 3–56.
- Frankel, R., and C. M. C. Lee. 1998. Accounting valuation, market expectation, and cross-sectional stock returns. *Journal of Accounting and Economics* 25: 283–319.
- Gow, I. D., G. Ormazabal, and D. Taylor. 2010. Correcting for cross-sectional and time-series dependence in accounting research. *The Accounting Review* 85 (2): 483–512.

- Jegadeesh, N., and S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance* 48 (1): 65–91.
- , J. Kim, S. D. Krische, and C. M. C. Lee. 2004. Analyzing the analysts: When do recommendations add value? *Journal of Finance* 59 (3): 1083–1124.
- Lakonishok, J., A. Shleifer, and R. W. Vishny. 1994. Contrarian investment, extrapolation, and risk. *The Journal of Finance* 49: 1541–1578.
- La Porta, R. 1996. Expectations and the cross-section of stock returns. *The Journal of Finance* 51: 1715–1742.
- Lee, C. M. C., and B. Swaminathan. 2000. Price momentum and trading volume. *The Journal of Finance* 55 (5): 2017–2070.
- Lin, H.-W., and M. F. McNichols. 1998. Underwriting relationships, analysts' earnings forecasts and investment recommendations. *Journal of Accounting and Economics* 25 (1): 101–127.
- Ljungqvist, A., C. Malloy, and F. Marston. 2009. Rewriting history. *The Journal of Finance* 64 (4): 1935–1960.
- Malmendier, U., and D. Shanthikumar. 2007. Are small investors naive about incentives? *Journal of Financial Economics* 85: 457–489.
- Miller, E. M. 1977. Risk, uncertainty, and divergence of opinion. *The Journal of Finance* 32: 1151–1168.
- Petersen, M. A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22 (1): 435–480.
- Piotroski, J. D. 2000. Value investing: The use of historical financial statement information to separate winners from losers. *Journal of Accounting Research* 38 (Supplement): 1–41.
- Pownall, G., and P. Simko. 2005. The information intermediary role of short sellers. *The Accounting Review* 80 (3): 941–966.
- Securities and Exchange Commission (SEC). 1999. *Short Sales*. SEC Concept Release No. 34-42037. File S7-24-99. Washington, D.C.: Government Printing Office.
- . 2008. Release No. 34-58591. Available at: <http://www.sec.gov/rules/other/2008/34-58591.pdf>.
- Seybert, N., and S. Wang. 2009. False prophets: How analysts lead short sellers astray during periods of high investor sentiment. Working paper. The University of Texas at Austin.
- Sloan, R. G. 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71 (3): 289–315.
- Swanson, E. P., L. Rees, and L. F. Juarez-Valdes. 2003. The contribution of fundamental analysis after a currency devaluation. *The Accounting Review* 78 (3): 875–903.
- Thompson Financial. 2009. *Methodology for Estimates*. Available at: http://www.library.kent.edu/files/TF_Methodology_for_Estimates_October_2009.pdf
- Womack, K. 1996. Do brokerage analysts' recommendations have investment value? *The Journal of Finance* 51: 137–167.

Incentive Compensation and Promotion-Based Incentives of Mid-Level Managers: Evidence from a Multinational Corporation

Merle Ederhof
University of Michigan

ABSTRACT: This study re-examines the hypothesis that explicit, compensation-based incentives of mid-level managers are adjusted to the level of implicit incentives provided by the possibility of moving to higher-level positions. Using compensation data from a large multinational corporation, I find that, after controlling for the position's scope and level of accountability, bonus-based incentives are stronger for managers who (1) have fewer organizational levels left to climb, (2) face weaker implicit incentives from getting promoted to the next level, and (3) face weaker implicit incentives from getting promoted to the top of the organization. The findings are consistent with the notion that implicit incentives are taken into consideration in the design of explicit incentive contracts. In particular, the results support the prediction that explicit incentives are optimally stronger in situations with weaker implicit incentives.

Keywords: *implicit incentives; incentive compensation; promotions; career concerns.*

Data Availability: *The data used in this study are proprietary.*

I. INTRODUCTION

The theoretical literature has argued that managers who face weaker promotion-based implicit incentives should optimally receive stronger explicit variable-pay-based incentives (Gibbons and Murphy 1992; Gibbs 1995). Despite well-developed theoretical arguments, empirical studies have had limited success providing evidence that implicit, promotion-based incentives are taken into consideration in the design of explicit incentive compensation contracts.¹ The present study revisits this hypothesis.

I thank Shannon Anderson (the editor), John Core, Ian Gow, Lisa Kahn, Brian Mittendorf, Madhav Rajan, Suraj Srinivasan, two anonymous reviewers, and seminar participants at the Yale Fall Conference, the NYU/Yale joint seminar, Michigan State University, University of Alberta, University of Michigan, University of Connecticut, and the AAA Annual Meeting. I also thank Bjoern Anton for excellent research assistance.

Editor's note: Accepted by Shannon Anderson.

¹ Gibbs (1995) analyzes the explicit incentives of employees who have been passed over for promotion; his findings do not indicate significant differences to the explicit incentives of employees who have not been passed over for promotion. Ortin-Angel and Salas-Fumas (1998) show that managers at higher organizational levels have stronger explicit incentives. However, their findings could be attributable to the fact that managers at higher levels are in positions that have more decision-making authority.

Submitted: June 2009
Accepted: June 2010
Published Online: January 2011

I analyze a sample of mid-level managers who can be directly compared with respect to their positions, but who occupy different positions in their respective hierarchies, and who face different promotion possibilities and rewards upon getting promoted. Thus, the setting provides an opportunity to observe variation in the strength of the implicit, promotion-based incentives while controlling for many confounding factors. In particular, the analyses in this study are based on compensation data from a large multinational corporation that operates in over 100 countries around the world. The company is organized around five main divisions, which in turn are comprised of 30 subdivisions. Each of the divisions and subdivisions is represented in many different countries. This matrix-like organizational structure allows me to directly compare positions across countries.

I find that the explicit incentives provided by the company's bonus plan are stronger for managers who are positioned at higher organizational levels, face weaker implicit incentives from getting promoted to the next level, and face weaker implicit incentives from getting promoted to the top of the organization, after controlling for the position's scope and level of accountability. These findings are consistent with the theoretical argument that implicit incentives should be taken into consideration in designing explicit incentive contracts. More precisely, the evidence presented here supports the prediction that explicit incentives are optimally stronger in situations where implicit incentives are weaker (Gibbons and Murphy 1992; Gibbs 1995).

This study makes several contributions to the incentive compensation literature. First, I extend prior empirical work in accounting that examines the incentive intensity of mid-level managers (Baiman et al. 1995; Nagar 2002) to consider the role of implicit promotion-based incentives, thereby following the call in Bushman and Smith (2001) for more research on the interactions between incentive contracting and other organizational features. In particular, the present study documents that the intensity of explicit incentives is higher in situations that pose weaker implicit, promotion-based incentives.

Second, this study contributes to prior studies that look at the interplay between explicit incentive contracts and implicit incentives arising from the possibility of career advancement (e.g., Kahn and Sherer 1990; Gibbons and Murphy 1992). Those studies primarily focus on the length of the manager's career horizon as the measure of the strength of implicit incentives. This study adds to that literature by analyzing an innovative setting that provides the opportunity to isolate the strength of implicit, promotion-based incentives. The analyses in this study provide additional empirical evidence on the importance of career-based incentives in contract design.

Arguably most closely related to this study, Gibbs (1995) analyzes the compensation-based incentives of employees who have been passed over for promotion. The author does not find significant differences between the explicit incentives of employees who have been passed over for promotion and those of employees who have not been passed over for promotion. Gibbs (1995) hypothesizes that the lack of evidence could be attributable to a centrally administered incentive scheme, which may not allow for variation at the employee level. In contrast to the setting analyzed by Gibbs (1995), the empirical setting in this study provides the following advantages. First, the data set used in this study includes explicit information on the parameters of the incentive contracts, such as the expected bonuses, which allows for a more precise measurement of the strength of the bonus-based incentives. Second, although Gibbs (1995) and the present study both analyze data from a single company, in my empirical setting, the explicit incentives are determined by the individual country organizations. Thus, my analysis circumvents the issue of a centrally administered incentive plan.

The ability to generalize the results of this study is limited by the analysis of a single firm. However, the research site chosen for this study offers several advantages for empirical investigation of the role of implicit, promotion-based incentives in incentive contracts provided to mid-level managers. First, the organizational structure of the company, combined with the company's

job-rating project, allows for a direct comparison of managers in different countries. Second, the compensation practices in the different countries reflect the local labor markets and, thus, are not specific to the company that is studied. Finally, the incentives provided to the managers through the company's bonus plan are not based on company-wide guidelines, but are decided by the individual country organizations.

The next section develops the hypotheses. Section III describes the research site, the sample and the measures used in the empirical analyses. Section IV discusses the research design and the empirical results. Section V provides a summary and conclusion.

II. HYPOTHESIS DEVELOPMENT

This study analyzes the incentives of mid-level managers who work in a corporate hierarchy. In particular, I investigate whether explicit incentives that are provided by variable-pay-based schemes are adjusted, based on the level of implicit incentives that are provided by the possibility of moving to higher-level positions in the organization.²

The argument that explicit, variable-pay-based incentives are optimally stronger in situations where implicit, promotion-based incentives are weaker has been formalized in career concerns models that allow for the presence of explicit incentive contracts (Gibbons and Murphy 1992; Gibbs 1995).

In particular, Gibbons and Murphy (1992) investigate optimal (explicit) incentive contracts when the agent faces implicit incentives from the possibility of career advancement in a competitive external labor market. The analysis in Gibbons and Murphy (1992) is based on a multiperiod model with a single performance measure, which is a function of the agent's innate ability and the agent's effort that is provided during the period. The performance metric is used in the explicit incentive contract and is also used by the external labor market to update beliefs about the agent's ability. The compensation that is offered to the agent in the second period in the competitive labor market is increasing in the market's assessment of his/her ability. Implicit incentives arise in this setting because ability and effort cannot be fully separated and the agent has the incentive to increase effort to influence the labor market's beliefs about his/her ability.³ The implicit incentives provided by career concerns are stronger when future compensation is more valuable to the agent as is the case when the agent is further away from retirement. Given this setup, the analysis shows that the optimal explicit incentives provided by the agent's compensation scheme are decreasing in the implicit incentives provided by career concerns.

Although Gibbons and Murphy (1992) examine implicit incentives that arise in a competitive external labor market, their analysis can also be interpreted in the light of an internal labor market, as it is discussed by the authors on pages 469–470. In particular, career concerns also arise in an internal labor market if the employee's supervisor cannot perfectly distinguish between the employee's ability and his/her effort. The authors argue that in an internal labor market setting explicit incentives should be strongest for workers with weak promotion-based incentives such as workers at the top of the corporate hierarchy.

Consistent with the result in Gibbons and Murphy (1992), Gibbs (1995), using a single-period model, shows that the explicit incentives provided by a compensation scheme are optimally stronger when the implicit incentives provided by the possibility of being promoted are weaker. In Gibbs' (1995) model, the promotion decision depends on the outcome of the performance measure, which again provides incentives for the agent to increase his/her effort.

² Promotion-based incentives are one source of implicit incentives. Other forms of implicit incentives include incentives that are based on non-contractible performance measures (see, e.g., Ederhof 2010).

³ In equilibrium, the market's conjecture about the worker's ability is correct but the agent will exert higher effort than in the absence of career concerns because the market discounts his/her effort.

An important feature of the models in Gibbons and Murphy (1992) and Gibbs (1995) is that the strength of the implicit incentives is not a choice variable for the principal. In other words, both models characterize how the optimal explicit incentive contract should be designed for a given level of implicit incentives. This is in contrast to Lazear and Rosen (1981) and Rosen (1986) who analyze how the principal should optimally choose the compensation structure across hierarchical levels in order to optimize the resulting implicit incentives. It seems reasonable to assume that, in my empirical setting, the implicit incentives provided by the possibility of promotion are determined exogenously with respect to the manager's compensation scheme. The company investigated in this study is organized along five main divisions with operations in many different countries. The purpose of the local units of a division in the different countries is to implement the global strategy of the division in the individual countries. The organizational structure of the local units is largely standardized. For example, all worldwide local units of a division are organized around a local unit manager whose authority and responsibilities follow worldwide guidelines. Thus, it seems unlikely that the company adapts its organizational form in a given country to the local compensation structure. In other words, it seems unlikely that the company structures its organizational form around the local labor market in order to optimize the implicit incentives resulting from the possibility of career advancement. Moreover, the compensation paid to the local managers is dictated by the local labor market conditions (Gibbs 1995).

Broadly speaking, the result of the analyses in Gibbons and Murphy (1992) and Gibbs (1995) is that optimal explicit incentives are decreasing in the strength of the implicit incentives provided by the possibility of career advancement. With respect to the setting of a corporate hierarchy, the strength of the implicit incentives is determined by the extent to which additional effort changes the probability of getting promoted and the "prize" that the manager is awarded upon promotion. The prize of getting promoted, in turn, is comprised of the immediate increase in compensation and the option value of being eligible for future rewards deriving from further promotions (Rosen 1986; Gibbs 1995).

An important determinant of the strength of the implicit incentives that a manager faces is his/her hierarchical position, because managers who are closer to the top of their organization have "truncated" promotion paths. Thus, high-rank managers are expected to have stronger explicit incentives. However, managers who are at higher organizational levels are likely to have stronger explicit incentives due to their job characteristics. In particular, managers at higher hierarchical levels are likely to have higher marginal productivities with respect to their effort (Baker and Hall 2004). Moreover, managers at high organizational ranks are likely to have more decision-making authority and to have a larger span of control (Prendergast 2002; Nagar 2002; Wulf 2007). In order to isolate the strength of implicit incentives, I use various control variables to capture variation in such job characteristics in the empirical analysis. The hypothesis can be stated as follows (expressed in the alternative form):

Hypothesis: The explicit incentives provided by the variable-pay scheme are decreasing in the strength of the promotion-based implicit incentives that the manager faces, *ceteris paribus*.

III. RESEARCH SETTING AND MEASURES

Research Site

I test the hypothesis using data from a large multinational engineering corporation that operates in approximately 100 countries. The company primarily sells technology to utility and industry customers. The company's operations are organized into five main divisions, which in turn are

comprised of 30 subdivisions.⁴ The divisions are the central building blocks of the organization; they run the business lines from R&D to sales and they have primary P&L responsibility. For each of the divisions and subdivisions, a manager is responsible for the unit's worldwide operations. For each country in which a unit operates, a local manager is responsible for the unit's local operations. In addition, the countries have some infrastructure in the form of a country management team and support functions such as human resources, finance, legal, and communication.

I test the hypothesis by comparing the incentives of the local managers in the different countries. In particular, the sample is comprised of all local managers who are in positions that are part of their local organizational "ladders," which go all the way to the top of the organizations. In other words, the sample includes managers who hold positions that make them eligible to eventually rise to the top of their respective organization. The sample excludes positions like IT and legal, which constitute support functions at the company and for which the promotion possibilities of the positions' holders are limited.⁵

The Company's Job-Rating Project

In 2005, with the help of a consulting firm, the company started a project of assigning numerical ratings to the top positions in the company. The company initiated the project for the following reasons. First, and important for this study, the company wanted to generate a picture of the organization's hierarchy that reflects managers' promotion paths and that is independent of existing job titles.⁶ In interviews with the author, the company contact emphasized that employees often have mental models of the company hierarchy that is based on job titles that do not correspond to the true hierarchy of the organization. The company hopes that the job-rating project will improve the promotion process, in that it will facilitate, for example, promoting people to jobs that have a higher rating but that have the identical title as the employee's current position.

Second, the company wanted to gain an understanding of how comparable positions are compensated in the different countries. In interviews with the author, the company contact emphasized that, to that end, a key aspect of the job-rating project was that the ratings were not influenced by the current position holder's compensation. Moreover, in order to facilitate comparability across countries, the ratings were assigned based on a standardized scheme, which the consulting company developed based on an initial pilot study.

The ratings are assigned to a position, and not to the person currently holding the position. Thus, the ratings are independent of the current manager's performance. The ratings are assigned based on a combination of factors primarily capturing the position's scope and level of accountability. In particular, a position's scope is captured by the number of employees that a manager oversees and the revenue figure for which the manager is responsible. The level of accountability is largely captured by the decision-rights that are in the manager's hand.

In interviews with the author, the company contact emphasized that promotions occur from one rating category to the next. In 2008, the company completed rating all positions that fall into the top ten rating categories. The position of the chief executive officer is assigned a rating of 1; an example of a job with a rating of 10 is a manager who is in charge of the operations of a local subdivision with revenue of \$22 million and 40 employees.

⁴ Subsequently, I collectively refer to divisions and subdivisions as "units."

⁵ The classification of positions into "organizational ladder" and "support functions" is based on extensive interviews with the company contact.

⁶ The notion that the organization's hierarchy is based on promotion paths is consistent, for example, with the way the hierarchy in Baker et al. (1994) and Gibbs (1995) is defined.

The Company's Incentive System

Aside from a fixed salary, all managers in the sample are eligible for a bonus payment. The company has guidelines that pertain to all worldwide participants in the bonus plan. In particular, the performance measures used in the bonus plan, the weight that is placed on firm-wide versus divisional measures, and aspects of the pay-performance relation must follow the company's guidelines. In contrast, and important for this study, the company does not have worldwide guidelines with respect to participants' expected bonuses—i.e., the compensation that is paid out in the form of a bonus when the performance meets expectations. In other words, the company does not have guidelines with respect to the percentage of a participant's cash compensation that is variable. The level of the expected bonus, which is typically expressed as a percentage of base salary, is determined by the respective country management. Some countries interpret the bonus plan as a guaranteed 13th-month salary for all participants. These countries are excluded from the analysis.

In addition to the bonus plan, the company awards stock options to the top employees of the organization who hold positions that are largely in the top seven rating categories. In contrast to the company's bonus plan, all aspects of the stock option plan are centrally administered by the company's headquarters and follow worldwide company guidelines. In awarding stock options in 2008, the company made use of the newly available job ratings. In particular, the number of options awarded to managers in a given job category is fixed for a given job-rating category.

Sample and Measures

The analyses in this study are based on a data set that contains information for 1,151 managers in 14 countries that have been assigned ratings between 2 and 10. The data are largely for the year 2008. Table 1 shows the distribution of the sample across different job-rating categories and countries.

Table 2 and Figure 1 summarize compensation levels across organizational levels for the different countries. Specifically, the table and graph show how the median total cash compensation, which is calculated by summing base salary and expected bonus, varies across job categories in the different countries.

The compensation figures are expressed relative to the median compensation levels at job-rating category 10 in the respective country, which are normalized to 1.00.⁷ As one would expect, for the most part, the values of the compensation ratios are increasing as managers move to higher-level job categories. Visual inspection of Figure 1 suggests that the overall pay structures can be characterized as convex, which is consistent with findings in prior studies and the prediction from tournament theory (e.g., Lambert et al. 1993; Rosen 1986). Table 2 and Figure 1 also indicate substantial variation in the pay structures across countries. For example, the ratio of median pay levels in category 5 to category 10 is approximately 2 in Germany, but more than 4 in the United Kingdom.

Explicit Incentives

Testing the hypothesis requires a measure of the explicit incentives provided by the compensation scheme. As discussed above, stock options are awarded uniformly across countries and the number of options awarded is fixed for a given job-rating category. Therefore, I focus attention on the company's bonus plan. Conceptually, the strength of the incentives provided by a bonus plan is reflected in how much the agent's compensation increases when s/he increases his/her effort. As discussed above, the company has worldwide guidelines with respect to certain parameters of the

⁷ Confidentiality reasons preclude me from reporting dollar amounts.

TABLE 1
Distribution of the Sample across Different Rating Categories and Countries

Country	Rating										Total
	1	2	3	4	5	6	7	8	9	10	
Sweden	0	1	0	1	5	9	15	40	78	99	248
Finland	0	0	1	0	0	5	8	12	34	37	97
Germany	0	0	1	0	1	3	9	30	32	18	94
Italy	0	0	1	1	3	5	10	21	23	28	92
Switzerland	0	0	0	1	3	6	6	24	42	92	174
Norway	0	0	0	0	2	4	3	9	15	18	51
Poland	0	0	0	0	1	0	0	6	5	11	23
Russia	0	0	0	0	1	1	3	6	5	12	28
Spain	0	0	0	0	2	2	1	5	12	21	43
United Kingdom	0	0	0	0	1	2	7	8	17	12	47
United States	0	0	0	0	4	17	26	40	63	74	224
Denmark	0	0	0	0	0	1	1	3	5	1	11
New Zealand	0	0	0	0	0	1	0	1	3	1	6
Turkey	0	0	0	0	0	1	1	4	2	5	13
Total	0	1	3	3	23	57	90	209	336	429	1,151

TABLE 2
Median Pay Levels across Job-Rating Categories for the Different Countries
 (expressed relative to the median pay level for category 10 in the respective country)^a

Job-Rating Category	Sweden	Finland	Germany	Italy	Switzerland	Norway	Poland
2	5.93	—	—	—	—	—	—
3	—	5.69	3.26	—	—	—	—
4	3.81	—	—	2.77	3.40	—	—
5	2.80	—	2.05	2.42	2.34	2.50	3.76
6	1.98	2.45	1.64	2.80	2.06	1.66	—
7	1.65	1.65	1.46	2.29	1.80	1.69	—
8	1.25	1.31	1.15	1.43	1.29	1.40	1.69
9	1.08	1.23	1.04	1.27	1.07	1.12	1.24
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Job-Rating Category	Russia	Spain	United Kingdom	United States	Denmark	New Zealand	Turkey
2	—	—	—	—	—	—	—
3	—	—	—	—	—	—	—
4	—	—	—	—	—	—	—
5	3.81	2.89	4.35	3.16	—	—	—
6	3.28	2.06	3.24	1.83	3.38	2.88	4.39
7	1.94	1.70	1.77	1.63	1.16	—	4.31
8	1.60	1.48	1.54	1.38	1.10	1.66	2.71
9	1.38	1.15	1.08	1.11	0.79	1.24	1.57
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00

^a Pay is the sum of the manager's base salary and his/her expected bonus. Expatriates are excluded from the statistics.

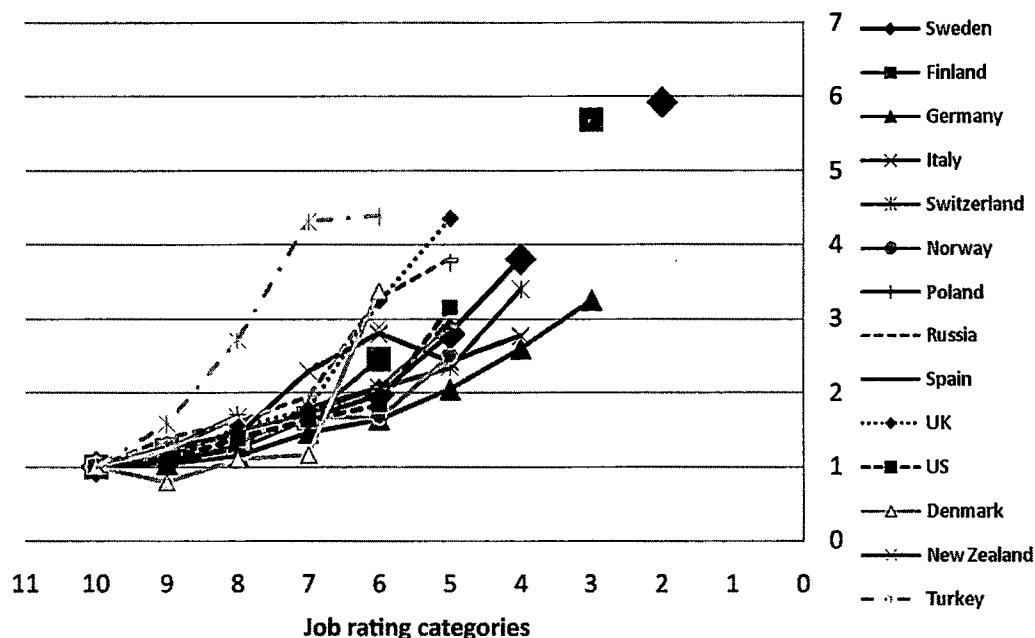
bonus plan. In particular, the payoff function is linear and an increase in performance leads to a fixed percentage increase in bonus payout. For example, the bonus payout for the maximum performance is twice as much as the bonus payout for the target performance for all plan participants. Therefore, variation in the expected bonus reflects variation in the strength of the incentives provided by the bonus plan. Thus, I measure the strength of the explicit incentives provided by the bonus scheme using the ratio of the target bonus to base salary (*TB*) (Indjejikian and Nanda 2002).

Implicit Incentives

The hypothesis predicts that managers who face stronger promotion-based implicit incentives have lower variable-pay-based explicit incentives. As discussed above, the strength of the implicit incentives is determined by the extent to which additional effort changes the probability of getting promoted and the "prize" that the manager is awarded upon promotion. The prize of getting promoted, in turn, is comprised of the immediate increase in compensation and the option value of being eligible for future rewards deriving from further promotions (Rosen 1986; Gibbs 1995). I use different measures to capture the strength of the implicit incentives.

First, as argued above, managers at higher hierarchical positions in their respective organizations have truncated promotion paths and are thus expected to have lower implicit incentives. In order to capture a manager's hierarchical position in his/her country organization, I include the job

FIGURE 1
(Median Cash Pay at Respective Level/Median Cash Pay at Level 10)^a



^a Expatriates are excluded from the statistics.

rating of the highest-ranking manager in the respective country, *HIGHESTPOSITION*, in the analysis.

As described above, the strength of the promotion-based implicit incentives is a function of the prize that the manager receives upon promotion and the extent to which additional effort changes the probability of getting promoted. I develop two additional measures to capture those features.

It is unobservable how additional effort changes the probability of getting promoted. However, Gibbs (1995, 1996) has shown that the derivative of the probability of getting promoted with respect to the agent's effort is increasing in the promotion probability as long as the promotion probability is below one-half (also see Campbell 2008). Interviews with the company contact confirmed that it can reasonably be assumed that, for the managers in my sample, the implicit incentives are increasing in the promotion probabilities, as the promotion rates at the company are sufficiently low.

Data limitations prevent the direct calculation of the promotion probabilities in the individual job categories. However, the company contact emphasized that dismissals and demotions are fairly rare in this company (also see Gibbs 1995). Thus, I employ the median tenure at the individual job levels as a proxy for promotion probabilities. More precisely, I use the inverse of the median tenure in the manager's current job-rating category to capture promotion possibilities.

Focusing on the manager's immediate promotion possibility, I construct a measure that takes the compensation differential between the manager's current job and the next-higher job-rating category, and the manager's chance of being promoted to that next level into consideration. Specifically, I compute the ratio of the median expected cash compensation for the next-higher job-rating category to manager's expected cash compensation (in his/her current job), both for the manager's respective country. I then multiply that ratio by the inverse of the median tenure in the manager's current job-rating category in the respective country. The measure is denoted by *IMPLICITNEXT*.

Similarly, consistent with Rosen (1986), I develop a measure that takes the median tenure and compensation level at the manager's current job and at each higher-level category in the respective organization into consideration. Specifically, I calculate the sum of the discounted compensation differentials between the manager's current job and the top of the respective organization. The compensation differential between two levels is discounted by the cumulative probability of being promoted to the respective level, where the probability of promotion is again proxied for by the inverse of the median tenure. I denote this measure as *IMPLICITTOP*.

Control Variables

Incentive compensation practices vary systematically across countries. As discussed in more detail in the research design section, I employ a measure that captures the general incentive intensity in the respective country. Specifically, I use a measure that is provided by the company's consulting company that captures the median incentive intensity for mid-level managers for a large number of industrial companies in the different countries.⁸

Prior literature has argued that division managers who have more decision-making authority receive more incentive-based pay because the potential for misuse is stronger if managers have more authority (Prendergast 2002; Nagar 2002; Wulf 2007). I use the job rating assigned to the manager's position, *CATEGORY*, as a proxy for the manager's decision-making authority. As described above, the ratings are assigned based on factors that primarily capture the positions' scope and level of accountability, an important determinant of which is the manager's level of decision-making authority.

A standard result in agency theory is that managers who have a higher marginal productivity with respect to effort should optimally have stronger explicit incentives (Baker and Hall 2004). In order to control for differences in managers' marginal productivities, following Wulf (2007), I include a measure of the relative importance of the manager's unit in the analyses. In particular, I measure a unit's relative importance by the ratio of the sales of the unit that the manager is affiliated with to the total sales in the respective country (*RSALES*).⁹

Theory suggests that more noise in the performance measures increases the risk that the manager is exposed to and the prediction is that incentives are lowered when the risk exposure is higher (Holmstrom 1979).¹⁰ As described above, the performance measures that are used in the bonus plans follow company guidelines in the sense that, worldwide, managers in the same position have the same performance measures in their bonus plan. Thus, managers in similar positions in different countries could be exposed to different levels of noise because they operate in different environments. Due to data limitations, I rely on the measure capturing the general incentive intensity in the different countries in order to capture differences in noise levels.

I also include the expected growth in sales in the regression analysis. I measure expected growth in sales by the ratio of budgeted sales for 2008 to the actual sales number for 2007

⁸ I do not report descriptive statistics on this measure due to confidentiality reasons.

⁹ Also see Baiman et al. (1995).

¹⁰ Also see Prendergast (2000) for arguments why the relationship between uncertainty and incentives could be positive.

(*SALESGROWTH*). This measure can be interpreted as a proxy for the firm's investment opportunities. Prior literature has argued that firms with greater growth opportunities employ compensation contracts with greater incentive intensity. Smith and Watts (1992) argue that the observability of managers' actions decreases with the firm's growth opportunities. In contrast, the actions of managers in low-growth firms are argued to be more observable because these actions are largely focused on the maintenance and supervision of existing assets (Gaver and Gaver 1993; Holthausen et al. 1995).¹¹

Agency theory has also argued that the optimal incentive intensity depends on the level of monitoring (e.g., Jensen and Meckling 1976; Prendergast 2002; Liang et al. 2008). In particular, Jensen and Meckling (1976) argue that incentive contracting and monitoring are alternative solutions to the moral hazard problem. As discussed above, the company is organized around five main divisions, which in turn are comprised of 30 subdivisions. The individual country organizations are structured around a country management team. Moreover, the local divisions and subdivisions are led by local division and subdivision managers. It seems reasonable to expect that direct monitoring of the actions of managers who are at the top of an organization is more difficult than monitoring the actions of lower-level managers. In order to capture such differences, I include an indicator variable, *TOPMANAGER*, indicating whether a manager is a highest-ranking manager in his/her country. Moreover, it seems plausible that local division managers do not receive as much monitoring as managers who are at lower organizational levels. Specifically, local division managers lead the respective business lines in their country. Thus, they are the highest-ranking managers in their respective fields of expertise. In order to capture potential differences in the monitoring of local division managers, I include an additional indicator variable, *DIVMANAGER*, in the tests. All measures are defined in Table 3.

Descriptive Statistics

Of the 1,151 managers in the sample, 18 are expatriates. In the company, expatriates are compensated based on the norms in their home country with certain adjustments, such as for hardship and allowance. Thus, expatriates are excluded from the analyses. With the exception of Italy, the expatriates hold lower-level positions, with ratings between 6 and 10. In Italy, the highest-ranking manager is an expatriate, which precludes calculation of the variable *IMPLICIT-TOP* for the observations from that country.

Table 4 provides descriptive statistics for the variables used in the analysis, calculating such statistics for each job-rating category using the pooled sample across all countries.

The mean (median) values for the dependent variable in the analyses, *TB*, which captures the ratio of (target bonus/salary), are 0.18 (0.14) for managers at level 10 and 0.39 (0.39) for managers who hold positions in job-rating category 3. Overall, the descriptive statistics indicate an increasing trend of *TB* as one moves to higher-level positions, which is consistent with the expectation that managers with more decision-making authority receive more incentive-based pay. However, the mean and median values for *TB* are fairly constant for job-rating categories 2 through 5. One issue that should be kept in mind is that the job-rating categories are not distributed evenly across countries and that there are country-specific differences in the level of incentive compensation. The finding that the mean and median values for *TB* are fairly constant for job-rating categories 2 through 5 could also be influenced by the fact that there is no one-to-one mapping between job-rating categories and hierarchical levels. Specifically, each of the categories 2 through 5

¹¹ Lambert and Larcker (1987) explain why the relationship between expected sales growth and the strength of the incentives could also be negative.

TABLE 3
Measures

Measure Reflecting the Strength of Explicit Incentives

TB_{ij} (target bonus/base salary)_{*i*}; both for manager *i*.

Measures Reflecting the Strength of the Implicit Promotion-Based Incentives

$HIGHESTPOSITION_j$ job-rating category of the highest-ranking manager in country *j*.

$IMPLICITNEXT_{ij}$ [(median expected cash pay_{*it-1*}/expected cash pay_{*i*}) * (1/median tenure_{*jt*})]: the ratio of the median cash pay (base salary + expected bonus) for job-rating category *t-1* in country *j* to the manager's expected cash pay in his/her current position.

$IMPLICITTOP_{ij}$ (median expected cash pay_{*jt-1*} / expected cash pay_{*i*})*(1 / median tenure_{*jt*})
 $+ \sum_{a=1}^{t-1} [(median expected cash pay_{ja-1} / median expected cash pay_{ja})$
 $* \prod_{b=0}^{t-a} (1 / median tenure_{ja+b})]$: sum of the discounted compensation differentials between the manager's current job-rating category *t* and the top of the respective country organization *j*.

Control Variables

$CATEGORY_{ij}$ job rating assigned to manager *i*'s position.

$RSALES_{ij}$ (sales of unit *k*/total sales of country *j* that unit *k* is located in).

$SALESGROWTH_{ij}$ (budgeted sales for 2008_{*k*}/actual sales for 2007_{*k*}); both for unit *k* in country *j*.

$TOPMANAGER_{ij}$ indicator variable that is equal to 1 if $CATEGORY$ indicates that the manager in job-rating category *t* is a highest-ranking manager in country *j*, 0 otherwise.

$DIVMANAGER_{ij}$ indicator variable that is equal to 1 if manager *i* is in charge of a local division, 0 otherwise.

includes a high percentage of managers who are highest-ranking in their respective country organization.

The descriptive statistics for *IMPLICITNEXT* suggest that the implicit incentives provided by the prospect of solely moving up one job category are largely increasing in the hierarchical level. Specifically, the mean (median) values for *IMPLICITNEXT* are 0.38 (0.37) for managers at level 10 and 0.61 (0.63) for managers who hold positions in job-rating category 5. This trend is consistent with the convex pay structures documented in Table 2 and Figure 1.

The variable *IMPLICITTOP* captures the implicit incentives provided by the possibility of moving to the top of the respective organization. Although the descriptive statistics largely indicate an upward trend with respect to the mean levels as one moves up the organizational hierarchy, they do not suggest a strong trend with respect to the median values. Specifically, with the exception of the figure for job-rating category 6, the median values of *IMPLICITTOP* are around 0.60 to 0.65 across categories. This finding is consistent with the notion in tournament theory that convex pay structures result in constant implicit incentives throughout the hierarchy (Rosen 1986). Broadly speaking, the intuition is as follows. On the one hand, managers who are further down in the hierarchy face a higher number of organizational levels that they could potentially climb, which increases the number of terms that comprise the option value, which

TABLE 4
Descriptive Statistics^a

	<u>n</u>	<u>Mean</u>	<u>Std.</u>	<u>Q1</u>	<u>Median</u>	<u>Q3</u>
CATEGORY = 2						
TB	1	0.38	NA	0.38	0.38	0.38
IMPLICITNEXT	NA	NA	NA	NA	NA	NA
IMPLICITTOP	NA	NA	NA	NA	NA	NA
TOPMANAGER	1	1.00	NA	1.00	1.00	1.00
DIVMANAGER	1	1.00	NA	1.00	1.00	1.00
RSALES	1	0.92	NA	0.92	0.92	0.92
SALESGROWTH	1	1.19	NA	1.19	1.19	1.19
CATEGORY = 3						
TB	2	0.39	0.02	0.38	0.39	0.40
IMPLICITNEXT	NA	NA	NA	NA	NA	NA
IMPLICITTOP	NA	NA	NA	NA	NA	NA
TOPMANAGER	2	1.00	NA	1.00	1.00	1.00
DIVMANAGER	2	1.00	0.00	1.00	1.00	1.00
RSALES	2	0.93	0.00	0.92	0.93	0.93
SALESGROWTH	2	1.15	0.01	1.14	1.15	1.16
CATEGORY = 4						
TB	3	0.38	0.14	0.25	0.38	0.53
IMPLICITNEXT	NA	NA	NA	NA	NA	NA
IMPLICITTOP	NA	NA	NA	NA	NA	NA
TOPMANAGER	3	0.33	0.58	0.00	0.00	1.00
DIVMANAGER	3	1.00	0.00	1.00	1.00	1.00
RSALES	3	0.40	0.44	0.13	0.16	0.92
SALESGROWTH	3	1.27	0.23	1.12	1.15	1.53
CATEGORY = 5						
TB	22	0.38	0.10	0.34	0.38	0.38
IMPLICITNEXT	3	0.61	0.12	0.48	0.63	0.72
IMPLICITTOP	3	0.61	0.12	0.48	0.63	0.72
TOPMANAGER	23	0.48	0.51	0.00	0.00	1.00
DIVMANAGER	23	0.22	0.42	0.00	0.00	0.00
RSALES	23	0.41	0.40	0.10	0.24	0.94
SALESGROWTH	23	1.15	0.18	1.08	1.10	1.19
CATEGORY = 6						
TB	52	0.36	0.25	0.23	0.30	0.40
IMPLICITNEXT	9	0.83	1.14	0.24	0.29	1.47
IMPLICITTOP	9	0.94	1.08	0.40	0.46	1.55
TOPMANAGER	55	0.05	0.23	0.00	0.00	0.00
DIVMANAGER	55	0.11	0.31	0.00	0.00	0.00
RSALES	43	0.36	0.37	0.10	0.16	0.67
SALESGROWTH	43	1.13	0.14	1.07	1.12	1.17
CATEGORY = 7						
TB	85	0.31	0.23	0.17	0.25	0.40
IMPLICITNEXT	18	0.73	0.55	0.33	0.60	0.86
IMPLICITTOP	18	1.08	1.77	0.44	0.60	0.86

(continued on next page)

TABLE 4 (continued)

	<u>n</u>	<u>Mean</u>	<u>Std.</u>	<u>Q1</u>	<u>Median</u>	<u>Q3</u>
<i>TOPMANAGER</i>	85	0.00	0.00	0.00	0.00	0.00
<i>DIVMANAGER</i>	85	0.05	0.21	0.00	0.00	0.00
<i>RSALES</i>	74	0.20	0.28	0.03	0.10	0.21
<i>SALESGROWTH</i>	74	1.26	0.86	1.06	1.12	1.19
<i>CATEGORY = 8</i>						
<i>TB</i>	203	0.26	0.20	0.15	0.21	0.29
<i>IMPLICITNEXT</i>	54	0.43	0.18	0.30	0.45	0.56
<i>IMPLICITTOP</i>	54	0.71	0.23	0.53	0.64	0.93
<i>TOPMANAGER</i>	208	0.00	0.00	0.00	0.00	0.00
<i>DIVMANAGER</i>	208	0.04	0.20	0.00	0.00	0.00
<i>RSALES</i>	174	0.25	0.33	0.04	0.12	0.26
<i>SALESGROWTH</i>	174	1.14	0.18	1.06	1.11	1.17
<i>CATEGORY = 9</i>						
<i>TB</i>	323	0.20	0.12	0.11	0.16	0.23
<i>IMPLICITNEXT</i>	105	0.41	0.15	0.28	0.41	0.49
<i>IMPLICITTOP</i>	105	0.64	0.15	0.53	0.65	0.71
<i>TOPMANAGER</i>	331	0.00	0.00	0.00	0.00	0.00
<i>DIVMANAGER</i>	331	0.01	0.09	0.00	0.00	0.00
<i>RSALES</i>	279	0.19	0.26	0.04	0.10	0.20
<i>SALESGROWTH</i>	279	1.13	0.19	1.04	1.12	1.18
<i>CATEGORY = 10</i>						
<i>TB</i>	404	0.18	0.09	0.11	0.14	0.20
<i>IMPLICITNEXT</i>	179	0.38	0.08	0.33	0.37	0.41
<i>IMPLICITTOP</i>	179	0.58	0.09	0.55	0.58	0.62
<i>TOPMANAGER</i>	425	0.00	0.00	0.00	0.00	0.00
<i>DIVMANAGER</i>	425	0.00	0.07	0.00	0.00	0.00
<i>RSALES</i>	334	0.26	0.34	0.03	0.12	0.26
<i>SALESGROWTH</i>	334	1.16	0.72	1.07	1.12	1.19

^a See Table 3 for variable definitions. The descriptive statistics are calculated for the pooled sample across countries, excluding 18 expatriate observations.

determines the implicit incentives. On the other hand, managers who are further down in the organization face lower compensation increases by moving up the initial levels, which are weighted most heavily in the computation of the option value.

The variable *TOPMANAGER* indicates whether a manager is among the highest-ranking managers in the respective country organization. Table 1 shows that the highest-level managers are concentrated in job-rating categories 2 through 6. Similarly, the descriptive statistics for *DIVMANAGER* indicate that the proportion of managers who are in charge of a local division is higher in the higher-level job-rating categories.

RSALES captures the relative importance of the division or subdivision with which a manager is affiliated, measured by the ratio of the unit's sales divided by the total sales of the respective country. As one would expect, the summary statistics for *RSALES* largely indicate that managers in lower job-rating categories are affiliated with smaller units. The median values for *RSALES* are

around 0.11 for categories 7 through 10, around 0.20 for categories 4 through 6, and over 0.90 for categories 2 and 3.

The median values for *SALESGROWTH*, which are measured by the ratio of budgeted sales for 2008 to actual sales for 2007 for the respective unit, largely indicate an upward trend as managers move to higher-level job categories. Managers in categories 2 through 4 are in charge of units that experience growth between 15 percent and 20 percent; in contrast, the expected unit growth for managers in categories 5 through 10 is around 11 percent. This finding is consistent with the notion that units that face higher growth opportunities are managed by higher-level managers who are delegated more decision-making authority.

IV. ANALYSES AND RESULTS

Strength of Implicit Incentives

The aim of this study is to analyze whether explicit incentives that are provided by variable-pay-based schemes are adjusted, based on the level of implicit incentives that are provided by the possibility of moving to a higher-level position in the hierarchy. The overall strategy that I use to address this question is to compare the explicit incentives of the local managers in the different countries. Given the organizational structure of the company, managers in the same job-rating category can reasonably be assumed to hold fairly comparable positions across countries. However, there is likely to be substantial variation in the strength of the implicit incentives that managers in different countries face due to the following reasons.

First, as can be observed from Table 1, the hierarchical structures differ across countries. Specifically, the countries vary with respect to the job category of the highest-ranking manager. In interviews with the author, the company contact emphasized that these differences are largely attributable to differences in the sizes of the units in the individual countries. Managers who have more organizational levels left “to climb” have stronger implicit incentives, all else equal.

Second, variation in the pay structures across job-rating categories and variation in the probabilities of getting promoted in the different countries can also result in differences in the implicit incentives that the managers face. Managers who face larger compensation increases upon being promoted and who have higher chances of getting promoted face stronger implicit incentives, *ceteris paribus*.

In order for the managers’ implicit incentives to be determined by features of their local country organizations, their career paths inside this company have to be confined to their respective countries. This assumption seems reasonable for this setting for the following reason. The company pursues a strategy of being “multi-domestic,” in the sense that the individual country organizations and their managers are expected to be very familiar with the local markets, governments, and infrastructures in order to be able to respond quickly to changes in the local environments. Part of this strategy is to employ local managers. Inspection of the data reveals that, aside from the 18 expatriates discussed above, the vast majority of managers are local. The company’s use of predominantly domestic managers is consistent with other examples in the literature (Brickley et al. 2009). Thus, the individual country organizations appear to be fairly segregated, making it reasonable to assume that the majority of the managers in the sample do not move between countries.

Table 5 reports the correlation among the variables used in the analysis. *TB*, which captures the strength of the explicit incentives, is significantly positively correlated with the job-rating category of the highest-ranking manager in the respective country, but is not significantly correlated with either of the measures capturing the strength of the implicit incentives (*IMPLICITNEXT*, *IMPLICITTOP*). With respect to these preliminary findings, one should keep in mind that it is likely that incentive compensation practices systematically vary across countries for unobserved reasons. As expected, there is a significant correlation between *TB* and measures that

TABLE 5
Correlations among the Variables^a

	TB	HIGHEST- POSITION	IMPLICIT- NEXT	IMPLICIT- TOP	CATEGORY	Log(RSALES)	SALES- GROWTH	TOP- MANAGER	DIV- MANAGER
TB									
HIGHESTPOSITION	0.46***								
IMPLICITNEXT	0.01	0.09*							
IMPLICITTOP	-0.03	0.09*	0.76***						
CATEGORY	-0.36***	-0.03	-0.32***	-0.22***					
Log(RSALES)	-0.03	-0.03	0.20***	0.07	-0.11***				
SALESGROWTH	-0.09**	-0.16***	0.04	0.03	-0.06*	0.04			
TOPMANAGER	0.23***	0.10***	NA	NA	-0.40***	0.16***	-0.01		
DIVMANAGER	0.19***	0.05	0.20***	0.27***	-0.32***	0.05	0.02	0.17***	

*, **, *** Significant at 10 percent, 5 percent, and 1 percent, respectively, all based on two-tailed tests.

^a See Table 3 for variable definitions. The statistics are calculated for the pooled sample across countries, excluding 18 expatriate observations.

indicate that the manager has a high rank in the respective organization (*CATEGORY*, *TOPMANAGER*, *DIVMANAGER*).

The two measures capturing the strength of the implicit incentives, *IMPLICITNEXT* and *IMPLICITTOP*, are highly correlated with each other; they are also significantly correlated with measures that indicate that the manager has a high rank in the respective organization (*CATEGORY*, *DIVMANAGER*). The latter finding is consistent with the picture that emerges from the descriptive statistics in Table 3. Namely, the average values for the strength of the implicit incentives are increasing as one moves up the organizational ladder.

The hypothesis predicts that the explicit incentives provided by the company's compensation scheme are decreasing in the strength of the promotion-based implicit incentives that the manager faces. In order to investigate the hypothesis, I employ *TB*, which captures the ratio of the expected bonus to base salary, as the dependent variable. As discussed above, I employ several measures to capture the strength of the implicit incentives that the manager faces. Specifically, the different measures capturing the strength of the implicit incentives are the job-rating category of the highest-ranking manager in a respective country (*HIGHESTPOSITION*) and measures that take compensation differentials between adjacent job-rating categories as well as promotion possibilities into consideration (*IMPLICITNEXT*, *IMPLICITTOP*).

As mentioned above, it is plausible that incentive compensation practices vary systematically across countries for unobserved reasons. In order to control for such country-specific differences, I subtract from *TB* a measure that captures the median incentive intensity for mid-level managers for a large number of companies in the respective country. Specifically, the measure, which was provided by the company's consulting company, captures the median incentive intensity for managers who hold comparable positions in industrial corporations. The managers hold positions that are comparable to jobs that are ranked at level 10 in the company that is studied here. The consulting company computed the median incentive intensity using data for 30 to 150 companies in a given country. In order to control for systematic differences in the level of growth in the different countries, I adjust the variable *SALESGROWTH* for the respective country's growth in GDP from 2007 to 2008.

Since the sample includes measures that vary at the country level but that are constant for all managers within a country as well as measures that vary across managers, the structure of the data set can be described as hierarchical, with managers nested in countries considered lower in the hierarchy than countries (Bryk and Raudenbush 1992). Specifically, the job rating for the highest-ranking manager in a country, which is captured by the variable *HIGHESTPOSITION*, varies only at the country level.¹² Expressed in terms of the manager and country level of analysis, the model is as follows:¹³

Level 1:

$$\begin{aligned} TB_{ij} = & \beta_{0j} + \beta_1(IMP\_ITNEXT_{ij} \text{ or } IMP\_ITTOP_{ij}) + \beta_2CATEGORY_{ij} \\ & + \beta_3Log(RSALES_{ij}) + \beta_4SALESGROWTH_{ij} + \beta_5TOPMANAGER_{ij} \\ & + \beta_6DIVMANAGER_{ij} + \varepsilon_{ij} \end{aligned}$$

for manager i in country j , where $\varepsilon_{ij} \sim N(0, \sigma^2)$.

Level 2:

$$\beta_{0j} = \gamma_0 + \gamma_1HIGHESTPOSITION_j + u_{0j}$$

¹² See Anderson et al. (2000) for an application of hierarchical linear models in the accounting literature.

¹³ Because the distribution of *RSALLES* is skewed, I use a log-transformation.

for country j , where $u_{0j} \sim N(0, \tau^2)$.

Combined in expressed form, the model is as follows:

$$\begin{aligned}
 TB_{ij} = & \gamma_0 + \gamma_1 HIGHESTPOSITION_j + \beta_1 (IMPLICITNEXT_{ij} / IMPLICITTOP_{ij}) \\
 & + \beta_2 CATEGORY_{ij} + \beta_3 \text{Log}(RSALES_{ij}) + \beta_4 SALES GROWTH_{ij} + \beta_5 TOPMANAGER_{ij} \\
 & + \beta_6 DIVMANAGER_{ij} + u_{0j} + \varepsilon_{ij}
 \end{aligned} \quad (1)$$

Table 6 shows the results of the estimation of Equation (1). I first estimate the model using the job-rating category of the highest-ranking manager in the respective country as the measure capturing the strength of the implicit incentives (Model I). The hypothesis predicts that the coefficient on *HIGHESTPOSITION*, β_1 , is significantly positive. After controlling for the manager's job-rating category, the managers' implicit incentives are expected to be lower when the highest-ranking manager in the respective country has a lower job-rating category. The results indicate that, consistent with the prediction, the strength of the explicit incentives provided by the company's bonus plan is higher for managers who have fewer organizational levels left to climb.

In particular, the coefficient on *HIGHESTPOSITION* has a value of 0.025 and is significant at the 5 percent level, indicating that a manager's expected bonus decreases, on average, by approximately 2.5 percent of salary for each hierarchical level that s/he climbs. For example, a manager who has three organizational levels left to climb has an expected bonus that is lower by 2.5 percent of salary than the expected bonus of a manager holding a similar position but who has only two organizational levels left to climb. Given that the mean (median) values of the salary-scaled expected bonus are 0.36 (0.30) for managers who are at hierarchical level 6, the difference appears to be economically meaningful.

Although the findings with respect to the variable *HIGHESTPOSITION* are consistent with the hypothesis, they may be attributable to alternative explanations. Specifically, it is possible that the performance measures used in the bonus plan better reflect the actions of the managers who are closer to the top of their organization. Moreover, it is conceivable that managers who have fewer organizational levels left to climb have more decision-making authority that is not fully captured by the manager's job-rating category.

Model II is estimated by using *IMPLICITNEXT*, which is intended to capture the implicit incentives that the manager faces when focusing on the next-higher job-rating category. I predict a negative coefficient on *IMPLICITNEXT*. The explicit incentives that are provided by the company's bonus plan are expected to be lower when the compensation differential and/or the probability of getting promoted, which are captured by *IMPLICITNEXT*, are higher. The results obtained from estimating Model II are consistent with this prediction. Specifically, the coefficient on *IMPLICITNEXT* has a value of -0.079 and is significant at the 5 percent level. For example, if the median tenure is shorter by one year for managers at level 9, the value of *IMPLICITNEXT* increases by 0.29, which translates into an expected bonus that is lower by 2.2 percent of salary, on average. Similarly, a 50 percent increase in the median expected cash pay at the next level for managers in category 9 translates into an expected bonus that is lower by 1.8 percent of salary, on average.

Model III is estimated by using *IMPLICITTOP*, which is intended to capture the implicit incentives that the manager faces when considering the entire organizational ladder. The prediction and findings are consistent with the findings for Model II. In particular, the coefficient on *IMPLICITTOP* has a value of -0.047 and is significant at the 5 percent level. For example, if the median tenure is shorter by one year for managers at level 9, the value of *IMPLICITTOP* increases by 0.43, which translates into an expected bonus that is lower by 2.0 percent of salary, on average. Similarly, a 50 percent increase in the median expected cash pay at the next level for managers in

TABLE 6
Effect of the Strength of Implicit Incentives on Explicit Incentives^a

$$TB_{ij} = \gamma_0 + \gamma_1 HIGHESTPOSITION_{ij} + \beta_1 (IMPLICITNEXT_{ij} \text{ or } IMPLICITTOP_{ij}) \\ + \beta_2 CATEGORY_{ij} + \beta_3 \text{Log}(RSALES_{ij}) + \beta_4 SALESGROWTH_{ij} \\ + \beta_5 TOPMANAGER_{ij} + \beta_6 DIVMANAGER_{ij} + u_{0j} + \varepsilon_{ij}$$

Variable ^b		Model I	Model II	Model III
		Coefficient (z-statistic)	Coefficient (z-statistic)	Coefficient (z-statistic)
Intercept		0.243 *** (4.40)	0.605 *** (2.93)	0.577 *** (2.75)
HIGHESTPOSITION	(+)	0.025 ** (2.10)		
IMPLICITNEXT	(-)		-0.079 ** (-2.45)	
IMPLICITTOP	(-)			-0.047 ** (-2.11)
CATEGORY	(-)	-0.033 *** (-5.85)	-0.049 *** (-4.42)	-0.046 *** (-4.30)
Log(RSALES)	(+)	-0.001 (-1.63)	-0.010 (-1.42)	-0.012 * (-1.73)
SALESGROWTH	(+/-)	-0.019 (-1.40)	-0.075 (-1.22)	-0.076 (-1.16)
TOPMANAGER	(+)	0.040 ** (2.37)	NA	NA
DIVMANAGER	(+)	0.026 (1.52)	0.172 (1.25)	0.183 (1.26)
S.D. of Intercept (u_0)		0.071 ***	0.148 ***	0.140 ***
Number of observations		877	257	257
Pseudo R ²		39.85%	22.65%	23.73%

*, **, *** Significant at 10 percent, 5 percent, and 1 percent, respectively, all based on two-tailed tests.

^a Expatriates are excluded from the regression models. Reported are the coefficients from the models with z-statistics in parentheses. The models are estimated using standard errors that are clustered by country.

^b TB is calculated by subtracting the median incentive intensity in the respective country from TB as defined in Table 3; $SALESGROWTH$ is calculated by subtracting the respective country's growth in GDP from $SALESGROWTH$ as defined in Table 3; $\text{Log}(RSALES)$ is natural logarithm of $RSALES$. The remaining variables are defined in Table 3.

category 9 translates into an expected bonus that is lower by 1.0 percent of salary, on average.

The findings with respect to the variables *IMPLICITNEXT* and *IMPLICITTOP* are consistent with the hypothesis. Nevertheless, it is possible that alternative explanations influence the findings. For example, it is possible that managers who face only small compensation increases upon promotion to the next level have relatively more decision-making authority.

With respect to the control variables in Table 6, the coefficient on the manager's job-rating category has a statistically significant value of around -0.03 to -0.04 across the three specifica-

tions. The findings indicate that a manager's expected bonus increases by 3–4 percent of salary, on average, when s/he moves to the next job-rating category. In Model I, the coefficient on the indicator variable, which indicates whether the manager is among the highest-ranking managers, is significantly positive, indicating that top-ranking managers receive higher explicit incentives. Specifically, the coefficient on *TOPMANAGER* has a value of 0.040, which suggests that top-ranking managers have expected bonus payments that are higher by 4 percent of salary, on average. The coefficients on the remaining control variables are insignificant at conventional levels. Specifically, the coefficient on *DIVMANAGER* is not significant at conventional levels. A potential explanation is that the effect is subsumed by the variables *TOPMANAGER* and *CATEGORY*.

In all three models, the standard deviation of the intercept is statistically significant. This finding indicates there is significant country-level variation around the intercept. Stated differently, there remains significant unexplained variation at the country-level.

It is conceivable that the results in Table 6 are driven by the number of organizational levels that the manager has left to climb. In other words, it is possible that the results in Models II and III with respect to the variables *IMPLICITNEXT* and *IMPLICITTOP* are driven by the manager's hierarchical level. In order to isolate the strength of the implicit incentives deriving from compensation increases and promotion probabilities, I re-estimate Models II and III from Table 6 after including the measure *HIGHESTPOSITION*, which controls for the number of organizational levels that the manager has left to climb. The results in Table 7 indicate that, even after controlling for the number of organizational levels that are left for the manager to climb, explicit incentives are stronger when promotion-based implicit incentives are weaker.

In summary, the results in Table 6 and Table 7 indicate that the explicit incentives provided by the company's bonus plan are higher when the manager has fewer organizational levels left to climb, when s/he faces lower implicit incentives from moving to the next organizational level, and when s/he faces lower implicit incentives from moving to the top of the organization. These findings support the hypothesis that explicit incentives provided by variable pay schemes are stronger when promotion-based implicit incentives are weaker. In a broader sense, the results are consistent with the notion that implicit promotion-based incentives are taken into consideration in designing explicit incentive contracts.

Robustness Tests

The inferences are robust to sensitivity tests. First, the results are robust to estimating an OLS regression on the pooled sample where the intercept is treated as non-random. Second, I repeat the analyses by including an indicator variable for each of the five different divisions. The results are virtually identical. Third, in order to control for the manager's career horizon (Gibbons and Murphy 1992), I include the manager's age as an independent variable in the regression analyses, which does not change the results. I also repeat the analyses using the ratio of the actual bonus that was paid out to base salary as the dependent variable in order to control for potential differences in the countries' bonus payout practices. The inferences remain unchanged. Moreover, the inferences are robust to including the unit's sales as a measure of the unit's size in the regression models. The results also remain unchanged when the variable capturing the manager's job-rating category, *CATEGORY*, is replaced by indicator variables in order to address potential nonlinearities.

V. CONCLUSION

This study re-examines the hypothesis that the explicit, compensation-based incentives of mid-level managers are adjusted to the level of implicit, promotion-based incentives. Specifically,

TABLE 7
Effect of the Strength of Implicit Incentives on Explicit Incentives^a

$$TB_{ij} = \gamma_0 + \gamma_1 HIGHESTPOSITION_j + \beta_1 (IMPLICITNEXT_{ij} \text{ or } IMPLICITTOP_{ij}) \\ + \beta_2 CATEGORY_{ij} + \beta_3 \text{Log}(RSALES_{ij}) + \beta_4 SALES GROWTH_{ij} \\ + \beta_5 TOPMANAGER_{ij} + \beta_6 DIVMANAGER_{ij} + u_{0j} + \varepsilon_{ij}$$

Variable ^b		I	II
		Coefficient (z-statistic)	Coefficient (z-statistic)
Intercept		0.420 *** (4.51)	0.402 ** (3.56)
<i>HIGHESTPOSITION</i>	(+)	0.031 (0.95)	0.031 (0.98)
<i>IMPLICITNEXT</i>	(-)	-0.086 *** (-4.30)	
<i>IMPLICITTOP</i>	(-)		-0.057 ** (-2.18)
<i>CATEGORY</i>	(-)	-0.046 *** (-5.01)	-0.045 *** (-4.62)
<i>Log(RSALES)</i>	(+)	-0.003 (-0.27)	-0.005 (-0.48)
<i>SALES GROWTH</i>	(+/-)	-0.060 (-0.92)	-0.065 (-0.93)
<i>TOPMANAGER</i>	(+)	NA	NA
<i>DIVMANAGER</i>	(+)	0.190 (1.28)	0.204 (1.31)
S.D. of Intercept (u_0)		0.130 ***	0.130 ***
Number of observations		257	257
Pseudo R ²		27.04%	28.27%

*, **, *** Significant at 10 percent, 5 percent, and 1 percent, respectively, all based on two-tailed tests.

^a Expatriates are excluded from the regression models. Reported are the coefficients from the models with z-statistics in parentheses. The models are estimated using standard errors that are clustered by country.

^b *TB* is calculated by subtracting the median incentive intensity in the respective country from *TB* as defined in Table 3; *SALES GROWTH* is calculated by subtracting the respective country's growth in GDP from *SALES GROWTH* as defined in Table 3; *Log(RSALES)* is natural logarithm of *RSALES*. The remaining variables are defined in Table 3.

this study revisits the theoretical argument that explicit incentives are optimally stronger in situations that pose weaker implicit, promotion-based incentives (Gibbons and Murphy 1992; Gibbs 1995).

The analyses in this study are based on compensation data from a large multinational corporation. This setting provides an opportunity to observe variation in the strength of implicit incentives because the sample is comprised of managers with comparable jobs but who face varying levels of implicit incentives since they are positioned at different hierarchical levels in their respective organization, their promotion possibilities vary, and they experience different levels of compensation increases upon promotion.

Regression analyses show that the incentives provided by the company's bonus plan are stronger for managers who are positioned at higher hierarchical levels, who face weaker implicit incentives from getting promoted to the next level, and who face weaker implicit incentives from getting promoted to the top of the organization, after controlling for the position's scope and level of accountability. These findings are consistent with the notion that implicit promotion-based incentives are taken into consideration in designing explicit incentive contracts, as proposed in the theoretical literature. More precisely, the evidence supports the hypothesis that explicit incentives are optimally stronger in situations with weaker implicit incentives (Gibbons and Murphy 1992; Gibbs 1995).

REFERENCES

- Anderson, S., D. Glenn, and K. Sedatole. 2000. Sourcing parts of complex products: Evidence on transactions costs, high-powered incentives and ex-post opportunism. *Accounting, Organizations and Society* 25 (8): 723–749.
- Baiman, S., D. Larcker, and M. Rajan. 1995. Organizational design for business units. *Journal of Accounting Research* 33 (Autumn): 593–616.
- Baker, G., M. Gibbs, and B. Holmstrom. 1994. The internal economics of the firm: Evidence from personnel data. *The Quarterly Journal of Economics* 109 (4): 881–919.
- , and B. Hall. 2004. CEO incentives and firm size. *Journal of Labor Economics* 22 (October): 767–798.
- Brickley, J., C. Smith, and J. Zimmerman. 2009. *Managerial Economics and Organizational Architecture*. New York, NY: McGraw-Hill Companies.
- Bryk, A. S., and S. W. Raudenbush. 1992. *Hierarchical Linear Models: Applications and Data Analysis*. Newbury Park, CA: Sage Publications.
- Bushman, R., and A. Smith. 2001. Financial accounting information and corporate governance. *Journal of Accounting and Economics* 32 (1–3): 237–333.
- Campbell, D. 2008. Nonfinancial performance measures and promotion-based incentives. *Journal of Accounting Research* 46 (2): 297–332.
- Ederhof, M. 2010. Discretion in bonus plans. *The Accounting Review* 85 (6): 1921–1949.
- Gaver, J., and K. Gaver. 1993. Additional evidence on the association between the investment opportunity set and corporate financing, dividend, and compensation policies. *Journal of Accounting and Economics* 16 (1–3): 125–160.
- Gibbons, R., and K. Murphy. 1992. Optimal incentive contracts in the presence of career concerns: Theory and evidence. *The Journal of Political Economy* 100 (June): 468–505.
- Gibbs, M. 1995. Incentive compensation in a corporate hierarchy. *Journal of Accounting and Economics* 19 (2–3): 247–277.
- . 1996. Promotions and incentives. Working paper, University of Chicago.
- Holmstrom, B. 1979. Moral hazard and observability. *The Bell Journal of Economics* 10 (Spring): 74–91.
- Holthausen, R., D. Larcker, and R. Sloan. 1995. Business unit innovation and the structure of executive compensation. *Journal of Accounting and Economics* 19 (2–3): 279–313.
- Indjejikian, R., and D. Nanda. 2002. Executive target bonuses and what they imply about performance standards. *The Accounting Review* 77 (4): 793–819.
- Jensen, M., and W. Meckling. 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3 (4): 305–360.
- Kahn, L., and P. Sherer. 1990. Contingent pay and managerial performance. *Industrial & Labor Relations Review* 43 (February): 107S–120S.
- Lambert, R., and D. Larcker. 1987. An analysis of the use of accounting and market measures of performance in executive compensation contracts. *Journal of Accounting Research* 25 (Supplement): 85–125.
- , ———, and K. Weigelt. 1993. The structure of organizational incentives. *Administrative Science Quarterly* 38 (September): 438–461.
- Lazear, E., and S. Rosen. 1981. Rank order tournaments as optimum labor contracts. *The Journal of Political Economy* 89 (October): 841–864.

- Liang, P., M. Rajan, and K. Ray. 2008. Optimal team size and monitoring in organizations. *The Accounting Review* 83 (3): 789–822.
- Nagar, V. 2002. Delegation and incentive compensation. *The Accounting Review* 77 (2): 379–395.
- Ortin-Angel, P., and V. Salas-Fumas. 1998. Agency-theory and internal-labor-market explanations of bonus payments: Empirical evidence from Spanish firms. *Journal of Economics & Management Strategy* 7 (Winter): 573–613.
- Prendergast, C. 2000. What trade-off of risk and incentives? *The American Economic Review* 90 (2): 421–425.
- . 2002. The tenuous trade-off between risk and incentives. *The Journal of Political Economy* 110 (October): 1071–1102.
- Rosen, S. 1986. Prizes and incentives in elimination tournaments. *The American Economic Review* 76 (September): 701–715.
- Smith, C., and R. Watts. 1992. The investment opportunity set and corporate financing, dividend, and compensation policies. *Journal of Financial Economics* 32 (3): 263–292.
- Wulf, J. 2007. Authority, risk, and performance incentives: Evidence from division manager positions inside firms. *The Journal of Industrial Economics* 55 (March): 169–196.

The Role of Organizational Absorptive Capacity in Strategic Use of Business Intelligence to Support Integrated Management Control Systems

Mohamed Z. Elbashir

The Australian National University

Philip A. Collier

The University of Melbourne

Steve G. Sutton

University of Central Florida

ABSTRACT: This study examines the influence of organizational controls related to knowledge management and resource development on assimilation (i.e., strategic integration and use) of business intelligence (BI) systems. BI systems use analytics and performance management concepts to leverage enterprise system databases and provide core management control system (MCS) capability. Our results indicate that organizational absorptive capacity (i.e., the ability to gather, absorb, and strategically leverage new external information) is critical to establishing appropriate technology infrastructure and to assimilating BI systems for organizational benefit. Further, findings show that while top management plays a significant role in effective deployment of BI systems, their impact is indirect and a function of operational managers' absorptive capacity. In particular, this indirect effect suggests that leveraging BI systems is driven from the bottom up as opposed to the top down. This differentiates BI from other isolated strategic MCS innovations that have traditionally been viewed as top-management driven.

Keywords: *business intelligence; management control systems; knowledge creation; absorptive capacity; management accounting systems; business analytics; corporate performance management; enterprise systems; enterprise resource planning systems.*

The authors thank Stewart Leech, Michael Davern, David Smith, Vicky Arnold, Shannon Anderson, Carlin Dowling, Uday Murthy, Anol Bhattacharjee, Andrea Kelton, Rod Dilnutt, Robyn King, and workshop participants at the Australian National University, University of Melbourne, University of Queensland, University of South Florida, and at the 2008 European Conference on Accounting Information Systems and the 2009 Accounting & Finance Association of Australia and New Zealand Annual Meeting for their valuable feedback at various stages of the research.

Editor's note: Accepted by James Hunton.

Submitted: April 2008

Accepted: June 2010

Published Online: January 2011

Data Availability: *Data are available through the first author, subject to limitations set by the Faculty Human Ethics Advisory Group and the vendor providing customer information.*

I. INTRODUCTION

Enterprise systems (i.e., integrated systems such as enterprise resource planning [ERP]) are fundamentally tied to the work of accounting. Enterprise systems have transformative implications for the integration of organizational information and control systems governing organizations (Chapman 2005; Chapman and Kihn 2009). Unfortunately, research indicates that most organizations' management control systems (MCS¹) have not leveraged this potential (e.g., Granlund and Malmi 2002; Dechow and Mouritsen 2005; Quattrone and Hopper 2005; Rom and Rohde 2007), but have simply used these systems to support existing MCS (Chapman and Kihn 2009). Enterprise systems are simply a resource that is made available and must be effectively leveraged for MCS enhancements to actually accrue (Chapman and Kihn 2009).

Business intelligence (BI) systems are widely viewed as the innovation that can leverage the wealth of data encapsulated in an enterprise system and support the anticipated transformation to a broader and more detailed MCS (Brignall and Ballantine 2004; Williams 2004; Carte et al. 2005; Kay 2006; Gnatovich 2007; Robertson et al. 2007). BI systems provide business analytics and corporate performance management reporting capabilities, the fundamental information for an effective MCS in a technology-driven business environment.² Davila and Foster (2007) identify 46 specific categories of MCS. The extensive set of pre-built reports and metrics included in most BI systems provide support for all of these categories.³ The purpose of this study is to examine the deployment of BI systems (an emerging MCS innovation) in order to better understand how an organization's controls related to knowledge management and resource development can lead to higher levels of BI assimilation.

Prior research suggests that integration of the enterprise system and MCS is heavily dependent on an organization's people, such that both the social and technical aspects of organizations should be considered in future research (Chapman and Kihn 2009). This view is consistent with Chenhall (2005), who argues that future research on MCS adoption needs to specifically consider how cultural controls influence implementation of an MCS innovation. Simon (1995, 34) defines value (i.e., cultural) controls as the "explicit set of organizational definitions that senior managers communicate formally and reinforce systematically to provide basic values, purpose, and direc-

¹ We refer to MCS as the formal, information-based routines and procedures that provide managers with measures, performance indicators, and procedures to maintain or alter patterns in the organizational activities to ensure that they are consistent with the organizational objectives and strategies (Simon 1995; Malmi and Brown 2008).

² Business Analytics provide the functionality to extract enterprise data and create a broad range of balanced scorecards, key performance indicators, total quality management metrics, and activity-based costing. Primarily they are associated with cybernetic controls while also providing some support for compensation/reward controls and administrative controls. Corporate Performance Management (CPM) capabilities provide support for budgeting, financial planning, consolidation, and enterprise planning. CPM also provides predefined reports for ABC, scorecards, and budgets. Thus, CPM primarily extends across planning and cybernetic controls (Howard 2003; Williams and Williams 2007).

³ The specific BI system used by the organizations that we examine includes three broad applications: financial, supply chain, and customer analyses. Financial analysis focuses on accounts receivable (customer credit, performance, cash inflow, and organizational effectiveness), general ledger (financial ratio reporting and analysis, financial performance, organizational performance, and budget analysis), and accounts payable (cash outflow, performance, vendor account analysis, and organizational effectiveness). Supply chain analysis focuses on inventory (stock overview and valuation, inventory forecasting, inventory demand, material movement activity, physical inventory, material reservations, and organizational effectiveness), procurement (material expenditure, procurement vendor analysis, material demand, process effectiveness, and organizational effectiveness) and the accounts payable analysis shared with the financial analysis application. Customer analysis includes sales (functional performance, customer sales, product sales, channel sales, distribution functional performance, sales organizational effectiveness, and distribution organizational effectiveness) and accounts receivable that is shared with financial analysis (Howard 2003).

tion.” Using that definition, we focus on one specific aspect of cultural control—that of control over knowledge management and resource development. The resulting culture determines the identification, absorption, and strategic application of new external information (hereafter, the level of *absorptive capacity*). Our focus includes the influence of the *top management team* (TMT) on the culture present at the operational manager level. By studying the level of absorptive capacity, we shed light on the interrelationship between cultural controls and deployment of an integrated MCS—i.e., MCS as a package (Malmi and Brown 2008).

The study of MCS as a package addresses another concern in the MCS innovation literature—the tendency to study MCS innovations in isolation. Chenhall (2003, 131) notes that MCS innovation studies are generally done in isolation. If specific accounting controls are systematically linked with other organizational controls, then research ignoring these connections may report spurious findings. BI systems provide a unique type of MCS innovation, in that they are not designed to support a single aspect of control (e.g., activity-based costing or balanced scorecards). Rather, BI systems provide powerful extraction capability to supplement substantial expansion of both planning and cybernetic controls, while also providing support for administrative and reward/compensation controls.⁴ In our study, we focus on organizations’ deployment of one BI system that is specifically designed to enhance MCS capability. At implementation, the system provides pre-developed metrics that can be configured to connect to the underlying databases of most leading ERP vendors. These metrics provide access to over 200 different pre-built reports using more than 500 key performance indicators and analytics answering over 2,900 business critical questions. The metrics include a broad array of performance measures for sales analysis, financial analysis, inventory and procurement analysis, and supply chain analysis—including a multitude of scorecard analyses (Howard 2003). Thus, amid concerns that MCS innovations have been studied in isolation from the overall package of MCS systems present in organizations (Chenhall 2005; Davila and Foster 2007; Malmi and Brown 2008), BI systems allow us to examine deployment of an integrated MCS.

We also address two general limitations that have been raised in regards to MCS adoption studies. First is the concern that studies have examined whether firms have adopted MCS, but those studies have not explored the variation in quality or depth of the use of MCS following that adoption (Davila and Foster 2007). Our research extends beyond adoption to assimilation, which includes the scope, use, and strategic integration of a system (Chatterjee et al. 2002). Second is the difficulty in moving beyond innovative case-based studies of MCS that can have limited generality (Davila and Foster 2005). Garnering access to multiple implementers of an innovation is generally quite difficult. Even when access may be possible, the variation in implementations often limits comparability. We avoided these problems by approaching one of the major BI vendors and securing their full client list. As a result, we acquire responses from a diverse sample of organizations implementing a common BI system.

⁴ Chenhall (2003) differentiates MCS from management accounting systems (MAS) by noting the latter is a subcomponent of the former. Chenhall (2003) defines the MAS as the systematic use of management accounting practices such as budgeting or product costing. MCS is broader and encompasses MAS as well as cultural and administrative controls. When we refer to core MCS capability, MAS capability is the primary interest that is encompassed primarily by cybernetic and planning controls (Malmi and Brown 2008). Malmi and Brown (2008) define the various categories of control as follows: *culture controls* establish the values, beliefs, and social norms that influence employees’ behavior; *cybernetic controls* provide quantitative measures for activities and processes, set performance standards, provide feedback processes, assess variance between goals and accomplishments, and influence individual behavior; *planning controls* set functional area goals, establish standards for assessing accomplishment, and assure alignment of goals across the functional areas of the organization; *reward/compensation controls* motivate and increase the performance of individuals; and *administrative controls* establish organizational structures, lines of accountability, and establish policies and procedures.

This study examines the role of top-management's and operational-level-management's absorptive capacity in facilitating assimilation of BI into the overall MCS. We develop an integrative model that theorizes relationships between the strategic and operational levels of absorptive capacity, and the impact on the existing IT infrastructure in order to improve our understanding of how organizations assimilate BI. We test the model using data collected from 347 business units that implemented the BI software. Our results indicate that organizational absorptive capacity is fundamental to both the readiness of the technology infrastructure for supporting BI integration and the successful assimilation into the overall MCS.

This study offers several contributions to the MCS literature. First, we examine the influence of the overall MCS on the adoption and strategic integration of BI, an MCS innovation. This is consistent with calls that advocate that MCS should be considered as a package, allowing MCS innovations to be examined in their context rather than in isolation. Second, in examining the cultural control aspect, we provide clarity to the role of TMT in assimilation of MCS innovation. By incorporating absorptive capacity, the result of cultural controls over knowledge management, we are able to show the critical intermediary role of operational management's absorptive capacity on assimilation of an MCS innovation. Third, our study moves beyond adoption of an MCS to consider the depth of strategic integration. By examining the assimilation of BI systems, we attain a better understanding of the strategic benefit that is derived from an MCS innovation. Fourth, we are able to establish that these relationships are consistent across a diverse range of organizations by examining one common BI system.

This study also contributes to the accounting information systems literature. First, for 30 years researchers have been focused on the normative development of integrated databases that unify enterprise-wide business event data (McCarthy 1982). This study provides empirical evidence of the benefit such integrated databases have for effective MCS. Second, our study provides insights into the strategic benefit that can accrue from standardized software extracting data from enterprise databases and providing both pre-specified and dynamic reporting capability. Third, we expand the general IT literature on the relationship between absorptive capacity and IT assimilation by focusing not only on the static knowledge component of absorptive capacity, but also the knowledge creation component. Our results indicate this latter component is critical to the assimilation of complex strategic systems such as BI.

Section II draws from theory on systems assimilation and the dynamic perspective of knowledge creation to develop a model of BI assimilation. Section III details the research methods, operationalization of constructs, and data collection process. Section IV details the results, and Section V discusses the implications.

II. THEORY AND HYPOTHESES

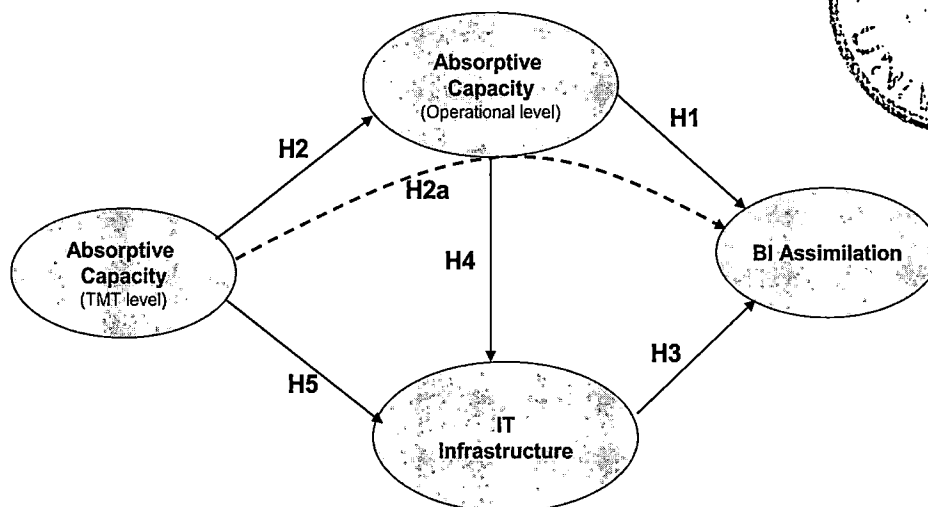
The adoption of enterprise systems and the resulting impact on MCS have been studied with great anticipation, given the potential for real benefits from integrated enterprise-wide data. Enterprise systems are expected to automate mundane MCS tasks and provide opportunity for broader-based MCS. These extended MCS are expected to enhance both management's strategic analyses and operational-level analyses as well as improve decision making (Sutton 2000; Granlund and Malmi 2002; Chapman 2005; Dechow and Mouritsen 2005; Quattrone and Hopper 2005; Arnold 2006). While this transformation is prevalent to some limited degree in some organizations (e.g., Granlund and Malmi 2002; Caglio 2003; Scapens and Jazayeri 2003; Quattrone and Hopper 2005), the general consensus is that little evolution in MCS has occurred. Rather, enterprise-wide data are often simply extracted for use by the same MCS modules that existed prior to implementation (Granlund and Malmi 2002; Dechow and Mouritsen 2005; Rikhardsson and Kraemmergaard 2006; Rom and Rohde 2007). Simply deploying enterprise systems appears insufficient to achieve significant enhancement of organizations' MCS (Chapman

and Kihn 2009). The professional literature similarly espouses the inadequacy of enterprise systems for facilitating the reporting and analyses necessary to support advanced MCS (e.g., Williams 2004; Kay 2006; Gnatovich 2007; Robertson et al. 2007; Williams 2008).

Enterprise systems provide highly integrative databases, but the ability for the average user to extract relevant data without the aid of specialized applications is limited. BI systems are designed to facilitate users in conducting detailed analyses of data contained in enterprise databases (Brignall and Ballantine 2004; Carte et al. 2005; Robertson et al. 2007; Williams 2008). BI systems hook into the underlying databases created by enterprise systems and provide a broad array of pre-specified reports and business analytics that form the information infrastructure necessary to support enhanced MCS capability.

By definition, assimilation studies are post-adoption studies that assume the decision to adopt a technology has been made, and acceptance and diffusion of the system is complete (Chatterjee et al. 2002). At an assimilation level, the interest is in whether a strategic technology has been integrated and used in a manner that provides strategic benefit, which is the focus of this study. Figure 1 provides an overview of our conceptual model. The model focuses on three critical organizational factors theorized to affect BI assimilation: TMT's absorptive capacity, operational managers' absorptive capacity, and sophistication of IT infrastructure. An organization's absorptive capacity is represented by its ability to recognize the value of new, external information, absorb it, and apply it for commercial ends (Cohen and Levinthal 1990). This focus on the absorptive capacity (Cohen and Levinthal 1990; Zahra and George 2002) of an organization's top-level and operational management is consistent with the views put forth by Chapman and Kihn (2009). This focus entails inclusion of the critical organizational components necessary for successfully leveraging enterprise systems. The joint effects of TMT and operational managers' absorptive capacity capture the effectiveness of cultural controls related to knowledge management.

FIGURE 1
Conceptual Model for BI Assimilation



Cultural Controls and the Development of Absorptive Capacity

Cultural controls relate to setting and instilling a set of values, beliefs, and social norms that are shared by an organization's members (Malmi and Brown 2008). Cultural controls encapsulate what Simons (1995, 34) identifies as "value" controls, and that he defines as "the explicit set of organizational definitions that senior managers communicate formally and reinforce systematically to provide basic values, purpose, and direction for the organization." Simons (1995, 34) notes further that these organizational definitions specify the "values and directions that senior managers want subordinates to adopt." Cultural controls play an important role in a broader view of MCS in that they set the boundaries and direction for core MCS capability (e.g., planning, cybernetic, and reward/compensation controls) and influence the role MCS plays in the organization (Chenhall 2003; Malmi and Brown 2008).

TMT represent a small group of the most influential executives with overall responsibility for an organization (Hambrick and Mason 1984). While MCS studies often highlight the critical role of TMT (Anderson and Young 1999; Bhimani 2003; Davila and Foster 2005; Naranjo-Gill and Hartman 2006, 2007; Davila and Foster 2007), TMT's knowledge has received limited attention in the MCS literature. TMT's knowledge is considered a primary indicator of TMT's competence (Armstrong and Sambamurthy 1999; Bassellier et al. 2003), yet research has failed to link TMT knowledge and assimilation (e.g., Armstrong and Sambamurthy 1999). We explore this phenomenon with the belief that prior studies, which have focused on a *static view*⁵ of TMT's knowledge, fail to capture the knowledge-creation activities that should drive technology innovation.

TMT knowledge from a *dynamic view*⁶ of organizational knowledge should provide a much better encapsulation of TMT's knowledge capabilities. In turn, a dynamic view should better capture the underlying theoretical reasons for the link between TMT and BI assimilation. An absorptive-capacity view focuses on the synergies of both TMT's knowledge and TMT's ability to put that knowledge into practice. From a theory perspective, TMT's absorptive capacity should be the key determinant of TMT's ability to provide effective leadership and support increased absorptive capacity at all levels of the organization. TMT's absorptive capacity is a key element for cultural controls to be effective in promoting organizational absorptive capacity.

Within an MCS context, operational managers can play a key role in driving the use of BI systems to support an organization's MCS needs. The effect of TMT's absorptive capacity on assimilation is therefore best conceptualized as indirect, mediated by operational management's absorptive capacity. Case-based evidence of the phenomenon can be interpreted from Caglio (2003), where the CFO became the driving force behind the design and implementation of the enterprise system. In the course of taking this lead, the CFO empowered the management accountants to be retrained to take on more of an analytic role for the TMT and a consulting role to a broad range of functional areas. Hence, the CFO's quick adaptation to the enterprise system and his rapid acquisition of relevant knowledge and capabilities allowed him to be an effective force in the implementation. However, the actual individuals implementing the changes that facilitated

⁵ The *static view* of organizational knowledge perceives organizations as a stock of knowledge that an organization possesses (Nonaka 1994; Nonaka and Takeuchi 1995; Grant 1996; Cook and Brown 1999). Most assimilation studies that have drawn on the knowledge-based view follow the static view and capture the amount of knowledge that people possess. Cook and Brown (1999), among many scholars, criticize research on organizational knowledge which follows the static view (Nonaka 1994; Kim 1998; Nonaka and Toyama 2003). Cook and Brown (1999) refer to this traditional understanding of organizational knowledge as the *epistemology of possession* because organizational knowledge is treated as something people possess.

⁶ The *dynamic perspective* of organizational knowledge is consistent with a large body of current knowledge management literature (Ditillo 2004; Vera-Munoz et al. 2006) that suggests understanding the organization's knowledge through its dynamic capabilities of knowledge creation rather than the stock of knowledge (static view) that organizations possess (Nonaka 1994; Grant 1996; Cook and Brown, 1999).

the transformation are the operational-level accountants acquiring the new knowledge to facilitate the transformation. The CFO was able to communicate his vision, providing a strong cultural control that helped transform behavior. This organization's experience is consistent with theorizations that organizations implementing controls promoting organizational absorptive capacity are better able to overcome knowledge barriers and effectively utilize new systems (Attewell 1992; Fichman and Kemerer 1997; Ravichandran 2005).

Cohen and Levinthal (1990), incorporating Attewell's (1992) views on organizational learning and innovation, posit that an organization's absorptive capacity facilitates learning and, in turn, drives innovation. As noted earlier, an organization's absorptive capacity is represented by its ability to recognize the value of new, external information, absorb it, and apply it for commercial ends (Cohen and Levinthal 1990).⁷ Cohen and Levinthal (1990) suggest that *prior relevant knowledge* and *intensity of effort* are critical elements for developing effective absorptive capacity. Intensity of effort represents the energy that members of an organization commit to solving problems and creating new knowledge.⁸

Absorptive Capacity and BI Assimilation

Successfully assimilating a BI system that automates MCS capabilities aligned with business strategies is dependent upon the development of relevant knowledge and skills (Fichman and Kemerer 1997; Armstrong and Sambamurthy 1999). The development of relevant knowledge and skills within an organization is highly dependent on the cultural controls management has put in place and supported (Malmi and Brown 2008). These cultural controls are used by organizations to integrate individuals' absorptive capabilities into the organization's routine and practice (Nonaka 1994; Kim 1998; Van den Bosch et al. 1999). Organizational absorptive capacity reflects the competence attributed to an organization's members (Szulanski 1996; Zahra and George 2002; Bassellier et al. 2003).

Two levels of organizational absorptive capacity are critical to BI assimilation. *TMT's absorptive capacity* represents the collective ability of TMT members to recognize the value of new information gathered from both internal and external sources, absorb it, and apply it to support their leadership role in strategic planning and control. TMT's absorptive capacity is determined by the broader knowledge and expertise that its members possess as well as knowledge-creation activities, including interaction with external constituents, such as competitors, customers, and peers (Nambisan et al. 1999; Daghfous 2004). *Operational-level managers' absorptive capacity* refers to the ability of managers at the operational level to value new information, absorb it, and apply it to support the organization's business strategy and value chain activities. Absorptive capacity at the operational level is heavily influenced by cultural controls, including TMT inter-

⁷ This definition of absorptive capacity suggests that the organization's absorptive capacity is built upon three capabilities: value, absorption, and application (Cohen and Levinthal 1990; Zahra and George 2002). (1) *Value capability* refers to the organization's ability to recognize new information, whether received from internal or external sources. This capability requires the organization to possess prior relevant knowledge and expertise that will help to value/assess new information. (2) *Absorption capability* refers to the organization's ability to analyze, process, interpret, and understand the new information. (3) *Application capability* refers to the ability of the organization to use the new information to support an organization's activities and strategies. The three capabilities are combinative in nature as they build upon each other to create the dynamic capability of the organization. This also suggests that an organization's absorptive capacity is path-dependent, as prior relevant knowledge must exist in order to facilitate absorption and use of new information.

⁸ Kim (1998) suggests a matrix that describes the interaction between prior relevant knowledge and intensity of effort and the effect of that interaction on the organizational absorptive capacity. According to Kim (1998) the absorptive capacity of the organization will be at its highest level when both the intensity of effort (i.e., action) and prior relevant knowledge are high. The absorptive capacity will be at the lowest level when both prior relevant knowledge and knowledge-related effort are low. In this study we use the underlying concept articulated by this matrix to operationalize the absorptive capacity construct tested in our research model.

vention and focused knowledge-creation activities (Boynton et al. 1994; Fichman and Kemerer 1997; Jansen et al. 2005).

BI assimilation is largely focused on the support and integration of core MCS control processes (e.g., planning and cybernetic) that are of primary concern at the operational or line levels. These core control processes are generally thought of as the management accounting system (Chenhall 2003; Malmi and Brown 2008). We see parallels in the few reported successful enterprise system integration studies. For instance, Caglio (2003) notes the empowerment of management accountants across operational and line functions with their ability to leverage enterprise-wide data for a broad range of functional areas. We posit that operational-level absorptive capacity is a critical precursor to BI assimilation, leading to H1:

H1: Operational-level managers' absorptive capacity will positively enhance organizations' BI assimilation.

As noted earlier, TMT plays a key role in establishing the cultural controls that motivate and enable the development of absorptive capacity. TMT leadership roles can be viewed from two perspectives: external and internal (Ulrich and Wiersema 1989; Kakabadse et al. 1995). External leadership roles include the ability to interact with the changing environment and interpret this into internal vision (Daft and Weick 1984; Hambrick 1995). Internal leadership roles include the design and management of employees' actions that enable realization of an organization's vision (Kakabadse et al. 1995; Anderson and Young 1999; Caglio 2003; Chenhall and Euske 2007). This includes provision of knowledge-creation mechanisms that address knowledge gaps at both the TMT and operational levels and are necessary to support new strategies and enabling technologies (Keen 1991; Nonaka et al. 1998; Caglio 2003).

Identifying and remediating knowledge gaps is critical, as BI assimilation is dependent on operational-level managers understanding the full potential of BI systems. This requires operational-level managers to raise their IT literacy to a level conducive with effective deployment (Rikhardsson and Kraemmergaard 2006). This dependency on operational-level learning suggests the relation between TMT's absorptive capacity and BI assimilation is mediated by operational-level absorptive capacity. The second hypothesis is stated as:

H2: TMT's absorptive capacity will positively enhance operational level-managers' absorptive capacity.

H2a: TMT's absorptive capacity will positively enhance organizations' BI assimilation, through the operational-level managers' absorptive capacity.

IT Infrastructure Sophistication and BI Assimilation

IT infrastructure sophistication refers to "the extent to which an organization has diffused the key information technologies into its foundation for supporting business applications" (Armstrong and Sambamurthy 1999, 309). IT infrastructure sophistication reflects the diversity and integration of IT components necessary to support BI systems. BI is designed to leverage complex business data (Quattrone and Hopper 2005; Dechow and Mouritsen 2005) that are integrated with other business information in the creation of enterprise-wide databases (Granlund and Malmi 2002). Accordingly, we posit that organizations with the sophisticated IT infrastructure to support information systems integration and enterprise-wide data integration will be better able to successfully assimilate BI systems. This reasoning leads to H3:

H3: The sophistication of IT infrastructure will positively enhance an organization's BI assimilation.

Absorptive Capacity and IT Infrastructure Sophistication

IT users at the operational level are important sources of IT innovation (von Hippel 1994; Nambisan et al. 1999). In an enterprise systems environment, operational-level managers are capable of driving systems configuration (Caglio 2003; Rikhardsson and Kraemmergaard 2006; Byrne and Pierce 2007; Elbashir and Williams 2007). The manner in which IT infrastructure is deployed can directly affect the information available to support MCS objectives (Quattrone and Hopper 2005). Supporting broad-based MCS capabilities necessitates sophisticated infrastructure configurations only possible through high levels of operational-level absorptive capacity (Dechow and Mouritsen 2005). This reasoning leads to H4:

H4: Operational-level managers' absorptive capacity will positively enhance the IT infrastructure sophistication of the organization.

TMT contributes to various IT infrastructure-related activities including project planning, resource allocation, and user problem solving. Building sophisticated IT infrastructure to support various applications requires a sound understanding of the organization's strategic objectives and the types of IT infrastructure services required to support those objectives (Broadbent et al. 1999). TMT with higher absorptive capacity will be better able to align organization-wide IT infrastructure investments with business strategies. This reasoning leads to the following hypothesis:

H5: TMT's absorptive capacity will positively enhance the IT infrastructure sophistication of the organization.

Control Variables

Prior studies suggest that two ancillary factors can influence assimilation: time since adoption (Anderson and Young 1999) and firm size (Davila and Foster 2005, 2007). Two proxies are used for firm size: Number of employees and gross revenue of the firm (Zhu and Kemerer 2002; Subramani 2004; Liang et al. 2007). These factors are modeled as control variables to isolate primary influences on BI assimilation.

III. RESEARCH METHOD

We use a field survey method to test the research hypotheses. The survey was distributed to Australian client organizations of a single international vendor providing BI software. The vendor provides a major BI system used internationally and that is specifically recognized for its packaged MCS capability. The system is configurable for leading ERP packages, enabling the BI system to directly access the underlying enterprise databases. There are over 200 different pre-defined standard reports using over 500 key performance indicators (KPIs) and containing analytics to answer over 2900 business-critical questions. The metrics and reports support MCS capabilities for sales analysis, financial analysis, inventory and procurement analysis, and supply chain analysis, including a broad range of scorecards for assessing performance (Howard 2003). Focusing on a single software vendor controls for variation that may occur from differences in MCS capability across BI software options.

We distributed surveys to 1,873 managers in 612 organizations that the BI vendor provided (subject to a written nondisclosure agreement) as a contact list of clients. Where possible, we selected multiple respondents for each organization from the vendor's contact list to include senior executives, operational managers, and IT users. For a "small organization" that only had a single contact person, the organization was selected if the contact person was a senior executive (e.g., chief executive officer [CEO], chief financial officer [CFO], or chief information officer [CIO]). A multiple-respondent strategy is preferred for the richness of the data, to mitigate bias, and to enhance accuracy (Sethi and King 1994; Huber and Power 1985).

The survey protocol followed the guidance of Dillman (2000). We mailed survey packets including a cover letter, survey, and pre-paid reply envelope to each selected contact. A first reminder was sent by email four weeks later to all the recipients. A second survey packet was sent four weeks after the email reminder to all nonrespondents. A final reminder was emailed two weeks later with a URL link to the web-based version of the survey. Online surveys are used in prior studies for both sole and supplementary survey methods (Dillman 2000). Our tests reveal no differences between paper and online responses.

An average of three respondents in each targeted organization received the survey. We asked respondents from organizations with multiple strategic business units (SBUs) to choose whether to answer the survey on behalf of either a SBU or the whole enterprise. We received a total of 436 responses from 229 organizations including 65 online responses. Due to missing data and selection of "No Basis for Answering" responses on the main study variables, 17 responses are deemed unusable. This resulted in 419 usable responses, for a usable response rate of 22 percent and 36 percent for individual and organization response, respectively.⁹ The responses included 135 choosing to identify and respond on behalf of their SBUs. Including responses for SBUs and responses for organizations as a whole, the final sample consists of 347 organizational units.¹⁰

The average age of respondents is 41.1 years, with 16.5 years of work experience. Respondents are 80 percent male and 20 percent female. Respondents classifying themselves as business executives/managers comprise 54 percent, with 46 percent IT executives/managers and 13 percent holding both business and IT jobs. Over half (54 percent) of respondents report five to eight years of BI systems experience, while 26 percent have over eight years of experience. Large organizations are the predominant respondents, with an average of 663 employees and gross revenue exceeding AUD\$2 billion. The sample is also diverse in industry representation, with the most common being manufacturing (65, 18.7 percent) and Retail/Wholesale/Distribution (50, 14.4 percent). Double-digit responses were also received in Banking/Finance/Insurance, Transport/Logistics, Media/Entertainment/Publishing, Healthcare, Telecom, Agricultural/Mining/Construction, and Consulting/Professional Services.

To test the consistency of the responses, we compute the correlation of multiple responses from the same organization on the main constructs (see Armstrong and Sambamurthy [1999] for details). All correlations between two or more respondents from the same organization on the main constructs are significant ($p < 0.01$). The results provide strong evidence of the consistency between responses from a single organization. As a consequence, we use the average scores from multiple respondents as the organizational response. For organizations with a single respondent, the individual response is used to represent the organization.¹¹

We perform three tests for common method bias: Harman's one-factor, partialling out a "marker variable," and partialling out a general factor score (Podsakoff and Organ 1986; Podsakoff et al. 2003).¹² The results indicate no significant common method variance that threatens the quality of the data.

⁹ In addition, 91 respondents reported that they were not the correct informants to answer the survey. Another 70 sent their apology by email and often quoted reasons for not responding to the survey (such as company policy).

¹⁰ Early and late responses were compared in paired samples of 150, 100, 50, and 40 using an ANOVA test for nonresponse bias. The results show no significant differences on any of the study variables, including demographic and control variables. There is no indication of any nonresponse bias.

¹¹ An ANOVA test shows no significant differences between the individual and averaged responses.

¹² Harman's one-factor test showed neither a single factor emerged from the exploratory factor analysis nor did one general factor account for the majority of the variance in the measurement items used in the model. In partialling out an unrelated "marker variable" as a surrogate for common variance, we used theoretically unrelated variables (respondents' age and years of experience) and examined the structural model both with and without the "marker variables." The results suggest that neither of the two marker variables is statistically significant. In performing the *partialling out a general factor score* test, we added the highest factor from the "unrotated" exploratory factor analysis test to the PLS model as a control variable on the dependent and mediating variables in the model. It is assumed that this factor contains

Operationalization of Constructs

BI Assimilation

BI Assimilation is measured by adapting Armstrong and Sambamurthy's (1999) instrument for IT assimilation. Building on Porter's (1985) value chain framework, Armstrong and Sambamurthy (1999) use 14 items to measure IT assimilation, six that capture IT assimilation in different business activities and eight items that measure IT assimilation in business generic strategies. Armstrong and Sambamurthy (1999) report three dimensions for IT assimilation: logistics activities, marketing activities, and business strategies. A comparison of factors in Armstrong and Sambamurthy's (1999) measure with those identified by the literature as representative of advanced MCS suggest that most attributes involving enterprise data are captured. One item, "managerial processes," is added based on BI literature and feedback in a focus group meeting. This process results in a total of 15 items to measure BI assimilation.

Exploratory factor analysis supports the use of only 14 items. We eliminate one item from the measurement list due to high cross-loadings (more than or closer to the threshold of 0.50).¹³ The result from the exploratory factor analysis shows the remaining 14 items load on three factors of BI assimilation. These factors are referred to as (1) *customer relations*, (2) *business operations*, and (3) *marketing and sales*. Each factor combines both generic strategies and business activities related to a specific business function.

Organizational Absorptive Capacity

The measures of absorptive capacity at the TMT and operational levels in this study reflect the two core elements of the absorptive capacity: *prior relevant knowledge* and *intensity of effort* (Cohen and Levinthal 1990; Kim 1998). We use the four modes of knowledge creation suggested by Nonaka (1994): socialization, externalization, combination, and internalization¹⁴ as the basis to measure *intensity of effort* for both operational and TMT levels of the organization. The four knowledge-creation modes represent the organization's effort and ability to create and absorb new information from internal and external sources, convert it into new usable knowledge, and apply it to support strategic planning and execution of business strategies (Nonaka 1994; Davenport and Prusak 1998).

Operational-Level Absorptive Capacity

The measure of *intensity of effort* at the operational level consists of 25 items divided between socialization (seven items), externalization (six items), combination (five items), and internalization (seven items).¹⁵ Confirmatory factor analysis supports the use of only 22 items to measure the

the best approximation of the common method variance (Podsakoff and Organ 1986; Podsakoff et al. 2003). The findings show the original results are not affected by the general factor included in the model.

¹³ The new item "managerial processes" was removed from the measure of BI assimilation as it cross-loaded on to two factors. One plausible explanation for the cross-loading is that the item refers to broad managerial activities that span the whole value chain. Note also that our three dimensions of BI assimilation do not exactly align with those of Armstrong and Sambamurthy (1999). Our focus is on the whole value chain, while Armstrong and Sambamurthy (1999) tested the assimilation of "IT in general." However, all of their items do load on the overall construct and the variance in sub-categorization is likely due to the diverse nature of the industries and the different type of IT used in our study.

¹⁴ Nonaka (1994) and Nonaka and Takeuchi (1995) suggest that organizational knowledge creation is captured by four modes of knowledge conversion: *socialization* (the process of creating tacit knowledge through shared experience), *externalization* (the process of converting tacit knowledge into explicit knowledge), *combination* (the process of creating explicit knowledge from explicit knowledge), and *internalization* (the process of converting explicit knowledge into tacit knowledge). Each of the four modes creates new knowledge of a specific type, and the dynamic between the four modes enables organizational knowledge creation.

¹⁵ The derived measure of intensity of effort is based on three prior studies (Nonaka et al. 1994; Becerra-Fernandez and Sabherwal 2001; Choi and Lee 2002) that used Nonaka's (1994) four modes of knowledge creation. However, the three studies were conducted in three different countries (Japan, U.S., and South Korea) and the measurement items used in

four modes of knowledge creation. We eliminate three items from the operational-level, knowledge-creation measure because they load below 0.50 on the relevant dimension (Hair et al. 1998). We then use measurement indices produced by Smart PLS to examine the properties of measures for the knowledge-creation modes.¹⁶ All four measures show satisfactory levels of internal consistency and convergent and discriminant validity.¹⁷ We derive four composite variables from average scores corresponding to the four modes to proxy for intensity of effort by averaging the respondents' scores.

The *prior relevant knowledge* component of absorptive capacity captures the shared knowledge between line and IT managers. Nelson and Coopridge (1996) define shared knowledge as the *understanding* and *appreciation* among IT and line managers for the technologies and processes that affect their mutual performance. We use the five items suggested by Nelson and Coopridge (1996) to measure shared knowledge between line and IT managers, which represent two types of measures, multiplicative and general.

(1) For *multiplicative* or interaction measures, respondents are asked to assess separately the role of IS and line managers for two characteristics: Items 1 and 2 capture understanding, while items 3 and 4 capture appreciation. Using the conceptualization of fit interaction (Venkatraman 1989; Nelson and Coopridge 1996), we operationalize the two concepts of understanding and appreciation by multiplying the two relevant items for each of the concepts (i.e., $\text{item1} \times \text{item2}$ and $\text{item3} \times \text{item4}$).

(2) For *general* measures, we ask respondents to rate the overall level of appreciation line managers and IS managers have for each other's accomplishments (item 5).

This procedure results in three items to capture shared knowledge (one general item and two from the outcome of the multiplication). The multiplicative measures provide stronger evidence of the validity of the measurement instrument than would be possible from only one type of measure (e.g., general; Nelson and Coopridge 1996). This is because the distribution of the final measure score depends on the extent to which two indicators agree with each other.

We derive the operational-level, absorptive-capacity score by creating the interaction term between the intensity of effort and shared knowledge variables. We use the two-step score construction procedure suggested by Chin et al. (2003) to estimate the composite variable of the interaction term because the measurement of one of the interaction variables (intensity of effort) is modeled formatively (Chin et al. 2003; Lu and Wang 2008).¹⁸

these three studies are not the same. One concern was that some items used could be driven by country-specific culture that may not be relevant to Australian organizations. To overcome this concern, we synthesized the knowledge-creation modes measures reported in the three papers. To improve measure validity, items in common to at least two of the three instruments were chosen for the initial measurement draft. The initial measurement draft was subjected to a number of pilot-tests including interviews with senior managers, feedback from academics, two focus group meetings, and a small survey. As a result of the pilot testing, seven additional items were added to the initial measurement list.

¹⁶ We use Smart PLS to examine the measurement properties of the two measures (i.e., at both operational and TMT levels) of Nonaka's (1994) knowledge-creation modes using the two-stage method. The eight knowledge-creation modes (4 modes to proxy for TMT's intensity of effort and 4 to proxy for operational managers' intensity of effort) were concurrently examined in one model. The results indicate that each of the measurement modes shows high convergent and discriminant validity.

¹⁷ The results indicate the following for all measures: composite reliability measures are above 0.70, the average variance extracted is above 0.50, items load higher on their associated factors than on any other factors, and the square root of AVE of all constructs is higher than its correlation with any other construct in the model.

¹⁸ As Chin et al. (2003, 11) explain, the formative indicators are not assumed to reflect the same underlying construct (i.e., can be independent of one another and measuring different factors). Therefore, the product indicators (in our model the product of multiplying the four scores for intensity of effort by the three scores for shared knowledge) will not necessarily tap into the same underlying interaction effect. Following the two-step procedure, the first step entails using

TMT's Absorptive Capacity

For the intensity of effort element of absorptive capacity, we develop draft measures for Nonaka's (1994) four knowledge-creation modes using prior literature on TMT (Hambrick 1981, 1995; Hurst et al. 1989; Kakabadse et al. 1995; Kippenberger 1997; Wu et al. 2002). We include 31 items in the initial measurement draft. We discussed the initial measurement items with two senior managers and a knowledge management consultant. We ask the three managers in three separate meetings (1) whether the items included in the list represent the knowledge-creation activities that usually take place at the TMT level, and (2) whether there are any additional knowledge-creation activities relating to the TMT that should be included. This process was followed by several pilot tests including feedback from academics, two focus group meetings, and a small survey. Feedback from participants during the pilot tests was analyzed and incorporated in the measurement draft when appropriate. We include 26 items in the final survey, representing the four modes: socialization (seven items), externalization (six items), combination (six items), and internalization (seven items).

Exploratory factor analysis supports only 19 of the items for measuring knowledge creation at the TMT level. We eliminate seven items because they load below 0.50 on the relevant factor (Hair et al. 1998). All measures of the four knowledge-creation modes show satisfactory levels of internal consistency and convergent and discriminant validity.

We create four composite variables representing the four knowledge-creation modes by averaging the respondents' scores for each mode.¹⁹ The prior relevant knowledge component of absorptive capacity is measured using three items suggested by Armstrong and Sambamurthy (1999) to capture TMT's strategic knowledge of IT. The results reported in Table 1 indicate that the measure of TMT's strategic knowledge has high internal consistency and convergent and discriminant validity. We again follow Chin et al.'s (2003) two-step approach to create the TMT absorptive capacity composite scores using TMT's strategic knowledge and intensity of effort composite variables.

IT Infrastructure Sophistication

We measure IT infrastructure sophistication using a scale adapted from Armstrong and Sambamurthy (1999), refining the scale by drawing on literature from BI, interviews with BI professionals, and meetings with two focus groups. We assess the level of *IT infrastructure sophistication* by asking respondents to indicate the extent to which their organization had *diffused* the ten IT infrastructure components (see Table 1). The final measurement instrument consists of ten key IT infrastructure components that capture two dimensions of enterprise systems related IT infrastructure (i.e., generic and specialized IT infrastructure) and combine to provide the basis for a shared data environment.

the formative indicators in conjunction with PLS to create the underlying construct scores for predictor and moderator variables. In step two we use the two composite variables to create a single interaction term representing the absorptive capacity variable.

¹⁹ We conducted sensitivity tests by creating composite variables for each knowledge-creation mode (at both operational and TMT level) using weighted averages whereby each item load is multiplied by the corresponding measure score. We added the outcomes and calculated the average as the composite variables (modes) score. Results remain the same. Measurement properties were discussed in footnote 14.

TABLE 1
Individual Item Loadings, Composite Reliability, Average Variance Extracted (AVE) Statistics
Panel A: Assimilation—Customer Relation: (Composite Reliability = 0.92, AVE = 0.61)

	<u>Loading</u>	<u>Standard Error</u>	<u>t-statistics^a</u>
ASA1: Customer services (e.g., improving customer satisfaction).	0.77	0.03	29.34
ASA2: Delivery of products/services.	0.77	0.03	26.14
AST3: Enhancing customer relations.	0.75	0.03	22.68
AST5: Enhancing existing products/services.	0.80	0.02	32.31
AST6: Providing value-added goods/services to customers.	0.85	0.02	49.37
AST7: Creating new products/services.	0.73	0.03	26.53

Panel B: Assimilation—Business Operations (Composite Reliability = 0.88, AVE = 0.60)

	<u>Loading</u>	<u>Standard Error</u>	<u>t-statistics^a</u>
ASA5: Supplier management (e.g., inbound logistics or purchasing).	0.83	0.02	48.18
ASA7: Manufacturing and/or internal operations.	0.74	0.04	21.08
AST1: Being a low-cost producer/provider.	0.69	0.04	17.58
AST2: Creating flexible manufacturing/operations processes.	0.76	0.03	27.40
AST4: Enhancing supplier relations.	0.83	0.02	48.07

Panel C: Assimilation—Marketing and Sales (Composite Reliability = 0.87, AVE = 0.64)

	<u>Loading</u>	<u>Standard Error</u>	<u>t-statistics^a</u>
ASA3: Marketing (e.g., targeting customers and tailoring offers).	0.84	0.02	35.85
ASA4: Sales (e.g., sales force automation, revenue management).	0.84	0.02	41.81
AST8: Entering new markets.	0.80	0.02	35.73

(continued on next page)

Panel D: Operational Managers' Shared Knowledge: (Composite Reliability = 0.95, AVE = 0.86)

	<u>Loading</u>	<u>Standard Error</u>	<u>t-statistics^a</u>
Shared1*2: Line managers (Information systems manager) understand the work environment (problems, tasks, roles, etc.) of the information systems managers (line managers).	0.89	0.02	60.36
Shared3*4: Line managers (information systems managers) appreciate the accomplishments of the information systems managers (line managers).	0.95	0.01	162.48
Shared5: Information systems managers and line managers appreciate each others' accomplishments.	0.94	0.01	131.57

Panel E: TMT Knowledge: (Composite Reliability = 0.92, AVE = 0.80)

	<u>Loading</u>	<u>Standard Error</u>	<u>t-statistics^a</u>
BEK1: The potential and limitations of the organization's...	0.89	0.01	61.84
BEK2: The potential and limitations of emerging information...	0.93	0.01	107.65
BEK3: How competitors are applying information technologies...	0.86	0.02	41.62

Panel F: Generic-IT Infrastructure Sophistication (Composite Reliability = 0.86, AVE = 0.52)

	<u>Loading</u>	<u>Standard Error</u>	<u>t-statistics^a</u>
ICT1: Mainframe/server systems.	0.62	0.05	11.74
ICT2: Database/ERP systems.	0.56	0.05	11.51
ICT7: Internal computer network.	0.71	0.04	16.08
ICT8: IT architecture and standards.	0.77	0.02	33.83
ICT9: Security and risk management policies.	0.84	0.02	44.20
ICT10: The latest back-end technology.	0.77	0.02	31.18

Panel G: Specialized-IT Infrastructure Sophistication (Composite Reliability = 0.87, AVE = 0.64)

	<u>Loading</u>	<u>Standard Error</u>	<u>t-statistics^a</u>
ICT3: Data warehouse/data marts.	0.86	0.02	50.74
ICT4: Extract, Transform, Load (ETL) tools.	0.88	0.01	59.02
ICT5: Analytical/reporting tools.	0.77	0.03	23.88
ICT6: IT that enable electronic access to external data.	0.66	0.04	15.50

(continued on next page)

Panel H: TMT-level Knowledge Creation Modes used (proxy for Intensity of Effort at the TMT Level): Formative Construct

	Indicator Weight	VIF	Standard Error	t-statistics
SOC-TMT: Socialization	0.17	1.868	0.033	5.13
EXT-TMT: Externalization	0.26	2.782	0.040	6.44
COM-TMT: Combination	0.28	2.313	0.038	7.38
INT-TMT: Internalization	0.47	2.447	0.041	11.47

Panel I: Operational-Level Knowledge Creation Modes (Proxy for Intensity of Effort at Operational Level): Formative Construct

	Indicator Weight	VIF	Standard Error	t-statistics
SOC-OP: Socialization	0.29	1.979	0.044	6.63
EXT-OP: Externalization	0.19	1.812	0.046	4.16
COM-OP: Combination	0.27	2.086	0.044	6.14
INT-OP: Internalization	0.45	2.482	0.049	9.13

^a All t-statistics reported in this table are statistically significant at $p < 0.001$.

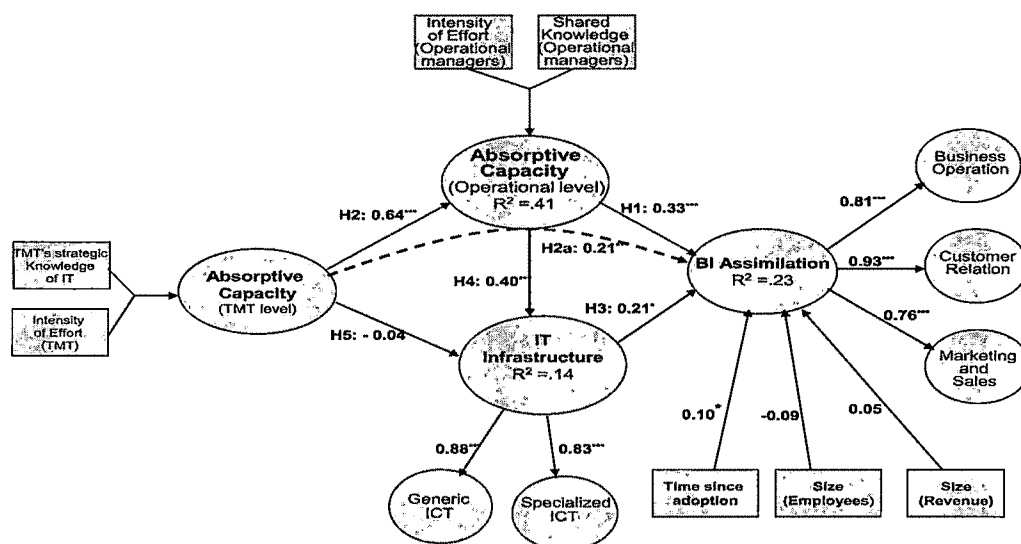
IV. RESULTS

We use partial least squares (PLS), a component-based structural equation modeling technique, to validate our constructs' measures and test the research model and hypotheses.²⁰ The overall results are shown on the structural model presented in Figure 2.

Measurement Properties

We apply multiple tests suggested by Churchill (1979) and Straub (1989) to assess construct validity and reliability. First, we use either *exploratory* or *confirmatory* factor analyses, depending on the maturity of the measure, to examine the dimensions and loading of measurement items. Second, we simultaneously test the structural model and the measurement model using PLS. The output from PLS in relation to the measurement model verifies the initial results from the exploratory factor analysis tests. Item loadings measure the significance of the item to the factor. For constructs measured with reflective indicators, we drop items with loadings below 0.70 from the construct measure as they indicate that less than 50 percent of the variance of the variable is

FIGURE 2
Structural Model Results for BI Assimilation



*, **, *** Indicate that the coefficient is significant at the $p < .05$, $p < .01$, and $p < .001$ levels, respectively.

²⁰ We use Smart PLS 2.0. PLS is a component-based structural equation modeling (SEM) technique that simultaneously tests the psychometric properties of the scales used to measure the constructs (i.e., measurement model) and examines the strength of the relations between the constructs (i.e., structural model) (Chin 1998; Hulland 1999). We use PLS instead of covariance-based SEM techniques (such as LISREL) because some of the constructs included in the research model are measured using formative indicators and using covariance-based modeling technique to test the research model can result in an unidentified model (Kline 2006). PLS is also suitable for analyzing complex models with mediating constructs and second-order constructs (Chin and Newsted 1999). A number of recent accounting research studies use PLS for similar reasons (e.g., Hall 2008; Dowling 2009; Chapman and Kihn 2009).

accounted for by the factor (Hair et al. 1998).²¹ However, we use a less strict rule of thumb for newly introduced measurement items (Chin 1998; Hulland 1999). To test formative construct validity we assess the indicators' weight rather than the loadings.

Construct Validity

Convergent Validity

Item loading, together with the average variance extracted (AVE), captures the convergent validity of each of the measures for constructs that are modeled reflectively (Van den Bosch 1999). Table 1 shows all reflective measurement items have high and significant loadings, and all the weights of items in formative constructs are statistically significant indicating their significant contribution to the measured construct. The AVE for all constructs exceed 0.50 (ranging between 0.52 and 0.86), supporting the convergent validity of the measurement items (Fornell and Larcker 1981).

Discriminant Validity

Table 2 shows that the values of the square roots of the AVE are all greater than the inter-construct correlations. This indicates that all measures have appropriate discriminant validity. An additional test of discriminant validity assesses each measurement item to ensure that it has a higher loading on its assigned factor than on the other factors (Chin 1998; Gefen et al. 2000) (see Table 3). Each measurement item loads higher on the appropriate construct than on any other construct (Chin 1998; Gefen et al. 2000), providing additional support as to the discriminant validity of the measures.

Hypotheses Testing

We use Smart PLS 2.0 with bootstrapping as a resampling technique (1000 random samples) to estimate the structural model and the significance of the paths. We use path coefficients and the R^2 jointly to evaluate the model (Chin 1998). Table 4 presents the PLS analysis results. As shown in Table 4, of the three control variables included in the research model (revenue and number of employees representing firm size in different ways, and time since adoption), only time since adoption is significantly related to BI assimilation. This indicates that organizations with more BI experience are able to assimilate BI better than those with less BI experience.

Overall, the results suggest the model has good predictability. The coefficients for all paths in the model, except one, are statistically significant at the 0.05 level. The results also indicate that 23 percent of BI assimilation, 41 percent of operational-level absorptive capacity, and 14 percent of IT infrastructure sophistication are explained by the model.

We test hypotheses within the structural equation model shown in Figure 2. Hypotheses that posit direct relations between constructs (H1, H2, H3, H4, and H5) were tested based on the magnitude and significance of path coefficients estimated using Smart PLS 2.0. We tested the hypothesis that posits a mediated (indirect) relation (H2a) by calculating the magnitude and sig-

²¹ The measurement items of BI assimilation, shared knowledge at operational level, and TMT's strategic knowledge, knowledge-creation modes (at both operational and TMT levels), and IT infrastructure sophistication are modeled reflectively, while the measurement items of intensity of effort (at both operational and TMT levels) are modeled formatively. The categorization of a construct as a formative or reflective is not always clear-cut and is influenced by the researcher judgment (Chwelos et al. 2001; Dowling 2009). We did additional tests similar to those reported in prior studies (Chwelos et al. 2001; Dowling 2009) whereby we examined different alternative measurement models, modeling all constructs in the model as formative, all the constructs as reflective, and various models with different constructs modeled as formative or reflective. All these tests show results similar to those reported in this paper as no paths gained or lost statistical significance or changed in sign. This suggests the results are not driven by how the constructs are modeled (i.e., formative or reflective).

TABLE 2
Inter-Construct Correlations and Square Root of Average Variance Extracted Statistics^a
(n = 347)

	Constructs Measured Using Reflective Items ^b							Constructs Measured Using Formative Items and Control Variables Measured with One Item				
	1	2	3	4	5	6	7	8	9	10	11	12
(1) Customer Relation (BI Assimilation)	0.79											
(2) Business Operations (BI Assimilation)	0.60	0.77										
(3) Marketing and Sales (BI Assimilation)	0.64	0.40	0.83									
(4) Generic ICT (IT Infrastructure Sophistication)	0.15	0.20	0.09	0.72								
(5) Specialized ICT (IT Infrastructure Sophistication)	0.42	0.29	0.29	0.47	0.80							
(6) Operational Managers' Shared Knowledge	0.25	0.22	0.30	0.24	0.18	0.93						
(7) TMT Knowledge	0.22	0.18	0.18	0.10	0.09	0.48	0.89					
(8) Intensity of Effort (Operational-Managers level)	0.37	0.41	0.33	0.37	0.33	0.54	0.44	1.00				
(9) Intensity of Effort (TMT level)	0.26	0.25	0.28	0.22	0.23	0.56	0.58	0.67	1.00			
(10) Firm Size (No. of Employees)	-0.08	-0.01	-0.05	0.17	0.17	-0.21	-0.18	-0.10	-0.15	1.00		
(11) Firm Size (Gross Revenue)	-0.03	0.02	0.04	0.12	0.17	-0.20	-0.19	-0.06	-0.07	0.67	1.00	
(12) Time Since Adoption	0.08	0.15	0.17	0.14	0.10	-0.03	0.07	0.15	0.06	0.16	0.22	1.00

^a Diagonal elements are the square roots of the average variance extracted statistics. Off-diagonal elements are the correlations between the latent variables calculated in PLS.

^b AVE will only be suitable to use when the construct is measured with reflective indicators (i.e., constructs 1 to 7).

TABLE 3
Items Loading and Cross Loading^a

Items	Constructs Measured Using Reflective Items						
	Customer Relation	Business Operation	Marketing and Sales	Generic Infrastructure	Specialized Infrastructure	Shared Knowledge (Operational)	TMT-Knowledge
ASA1	0.77	0.45	0.47	0.17	0.30	0.18	0.17
ASA2	0.77	0.56	0.38	0.15	0.33	0.12	0.17
AST3	0.75	0.46	0.53	0.10	0.34	0.19	0.08
AST5	0.80	0.44	0.51	0.10	0.36	0.16	0.15
AST6	0.85	0.48	0.51	0.09	0.40	0.30	0.19
AST7	0.73	0.43	0.58	0.11	0.24	0.22	0.27
ASA5	0.49	0.83	0.30	0.16	0.17	0.12	0.17
ASA7	0.34	0.74	0.29	0.13	0.12	0.15	0.12
AST1	0.38	0.69	0.20	0.24	0.29	0.19	0.11
AST2	0.48	0.76	0.39	0.13	0.28	0.21	0.14
AST4	0.58	0.83	0.36	0.13	0.25	0.18	0.14
ASA3	0.52	0.26	0.84	0.05	0.25	0.18	0.09
ASA4	0.51	0.32	0.84	0.09	0.28	0.26	0.12
AST8	0.55	0.41	0.80	0.09	0.19	0.30	0.22
ICT1	0.05	0.18	0.05	0.62	0.22	0.11	0.03
ICT2	0.08	0.21	0.04	0.56	0.34	0.05	0.06
ICT7	0.06	0.02	0.07	0.71	0.25	0.10	-0.03
ICT8	0.14	0.13	0.09	0.77	0.38	0.22	0.10
ICT9	0.09	0.16	0.04	0.84	0.36	0.22	0.13
ICT10	0.22	0.17	0.10	0.77	0.45	0.27	0.13
ICT3	0.36	0.23	0.22	0.38	0.86	0.14	0.12
ICT4	0.35	0.23	0.20	0.38	0.88	0.16	0.12
ICT5	0.37	0.18	0.32	0.31	0.77	0.15	0.03
ICT6	0.25	0.27	0.20	0.43	0.66	0.11	0.00
Shared1*2	0.21	0.18	0.30	0.25	0.19	0.89	0.46
Shared3*4	0.25	0.22	0.27	0.20	0.15	0.95	0.44
Shared5	0.25	0.21	0.26	0.22	0.16	0.94	0.43
BEK1	0.15	0.13	0.15	0.11	0.04	0.42	0.89
BEK2	0.15	0.16	0.15	0.07	0.10	0.45	0.93
BEK3	0.29	0.19	0.17	0.10	0.09	0.41	0.86
SOCO	0.35	0.39	0.23	0.32	0.27	0.42	0.33
EXTO	0.38	0.41	0.33	0.25	0.32	0.34	0.35
COMO	0.31	0.36	0.28	0.32	0.33	0.41	0.33
INTO	0.23	0.27	0.28	0.32	0.22	0.52	0.42
SOCs	0.24	0.26	0.22	0.21	0.21	0.36	0.35
EXTS	0.20	0.20	0.21	0.24	0.15	0.53	0.48
COMS	0.32	0.27	0.38	0.19	0.29	0.47	0.47
INTS	0.17	0.16	0.17	0.14	0.15	0.50	0.58
FIRMSIZE	-0.03	0.02	0.04	0.12	0.17	-0.20	-0.19
FIRMSIZEM	-0.08	-0.01	-0.05	0.17	0.17	-0.21	-0.18
YEAR	0.08	0.15	0.17	0.14	0.10	-0.03	0.07
Composite Reliability	0.92	0.88	0.87	0.86	0.87	0.95	0.92
AVE	0.61	0.60	0.64	0.52	0.64	0.86	0.82

(continued on next page)

**Constructs Measured Using Formative Items OR with One Item
(Control Variables)**

Items	Intensity of Effort (Operational)	Intensity of Effort (TMT)	Size (Revenue)	Size (Employees)	Time Since Adoption
ASA1	0.30	0.17	-0.03	-0.06	0.09
ASA2	0.22	0.16	-0.06	-0.07	0.13
AST3	0.26	0.17	0.06	-0.03	0.05
AST5	0.26	0.20	-0.03	-0.04	0.04
AST6	0.40	0.32	-0.04	-0.12	0.04
AST7	0.35	0.26	-0.04	-0.04	0.04
ASA5	0.34	0.21	-0.02	-0.04	0.14
ASA7	0.30	0.10	-0.05	-0.04	0.14
AST1	0.33	0.24	0.11	0.12	0.12
AST2	0.33	0.24	0.02	-0.02	0.06
AST4	0.36	0.21	0.02	-0.03	0.11
ASA3	0.22	0.18	0.07	0.01	0.14
ASA4	0.26	0.22	0.09	-0.02	0.11
AST8	0.35	0.31	-0.05	-0.11	0.17
ICT1	0.10	0.04	0.08	0.06	0.11
ICT2	0.18	0.08	0.16	0.19	0.14
ICT7	0.21	0.16	0.05	0.12	0.03
ICT8	0.36	0.21	0.15	0.18	0.13
ICT9	0.36	0.26	0.07	0.08	0.13
ICT10	0.31	0.19	0.02	0.09	0.06
ICT3	0.32	0.25	0.18	0.16	0.12
ICT4	0.29	0.22	0.16	0.18	0.10
ICT5	0.29	0.19	0.09	0.08	0.03
ICT6	0.21	0.08	0.12	0.13	0.07
Shared1*2	0.50	0.50	-0.17	-0.20	0.02
Shared3*4	0.48	0.54	-0.20	-0.20	-0.04
Shared5	0.48	0.51	-0.19	-0.19	-0.06
BEK1	0.37	0.51	-0.13	-0.16	0.08
BEK2	0.40	0.50	-0.19	-0.19	0.05
BEK3	0.42	0.51	-0.18	-0.13	0.05
SOCO	0.81	0.52	0.00	-0.05	0.17
EXTO	0.79	0.47	0.00	0.00	0.15
COMO	0.82	0.51	-0.08	-0.07	0.13
INTO	0.84	0.65	-0.09	-0.14	0.08
SOCs	0.52	0.77	0.01	-0.05	0.07
EXTS	0.54	0.88	-0.10	-0.17	0.07
COMS	0.58	0.84	-0.06	-0.09	0.07
INTS	0.57	0.87	-0.06	-0.16	0.02
FIRMSIZRE	-0.05	-0.06	1.00	0.67	0.22
FIRMSIZEM	-0.09	-0.14	0.67	1.00	0.16
YEAR	0.16	0.06	0.22	0.16	1.00

(continued on next page)

Constructs Measured Using Formative Items OR with One Item (Control Variables)					
Items	Intensity of Effort (Operational)	Intensity of Effort (TMT)	Size (Revenue)	Size (Employees)	Time Since Adoption
Composite Reliability	NA ^b	NA ^b	NA ^c	NA ^c	NA ^c
AVE	NA ^b	NA ^b	NA ^c	NA ^c	NA ^c

^a Cross-loading is obtained by calculating the correlation of the standardized latent variable scores with the standardized value of the item (Pavlou and Gefen 2005).

^b Composite reliability and AVE will only be suitable to use when the construct is measured with reflective indicators.

^c Control variables that are measured with one item.

TABLE 4
Path Coefficients: Test and Control Variables

	Path	Path Coefficient	t-statistic (z-score)
Hypotheses Direct Effect	Operational-level AbsCapacity → Assimilation of BI systems (H1)	0.33	6.37***
	TMT's AbsCapacity → Operational-level AbsCapacity (H2)	0.64	18.47***
	IT Infrastructure Sophistication → Assimilation of BI (H3)	0.21	3.47***
	Operational-level absorptive capacity ... IT Infrastructure Sophistication (H4)	0.40	6.87***
	TMT's AbsCapacity → IT Infrastructure Sophistication (H5)	-0.04	0.53
Hypotheses Indirect Effect	TMT'S AbsCapacity → Operational-level AbsCapacity → BI assimilation (H2a)	0.21 ^a	(6.10***)
	TMT'S AbsCapacity → Operational-level AbsCapacity → IT Infrastructure Sophistication (<i>ad hoc</i> test for H5)	0.25 ^a	(6.43***)
Control Variables	Size (No. of employees) → BI assimilation	-0.09	1.61
	Size (Revenue) → BI assimilation	0.05	0.89
	Time since adoption of BI → BI assimilation	0.10	1.98*

* t-statistic > 1.96 is significant at $p < 0.05$ level.

** t-statistic > 2.58 is significant at $p < 0.01$ level.

*** t-statistic > 3.29 is significant at $p < 0.001$ level.

^a Standardized path.

nificance of mediated paths (Hoyle and Kenny 1999; Subramani 2004). Additional tests examine whether the indirect path represents full or partial mediation (Baron and Kenny 1986; Mathieson et al. 2001).

Hypothesis 1 predicts that increased levels of operational-level absorptive capacity will enhance the levels of BI assimilation. The results shown in Figure 2 and Table 4 support the hypothesis with a strong and significant relationship (0.33, $p < 0.001$). This result supports the belief that operational and line managers' knowledge and knowledge-creation activities (i.e., absorptive capacity) have a positive influence on assimilation.

Hypothesis 2 predicts that increased levels of TMT absorptive capacity enhance the operational-level absorptive capacity. The results shown in Figure 2 and Table 4 again support the hypothesized relationship with a strong and significant relationship (0.64, $p < 0.001$). This finding

provides support for the integrative effects of cultural controls that primarily emanate from TMT-level management and the subsequent impact on organizational learning and knowledge creation at the operational and line levels.

Hypothesis 2a elaborates on this relationship to examine the connection that occurs between the outcome of cultural controls and the core MCS controls provided by BI systems (e.g., planning and cybernetic controls). Hypothesis 2a examines supporting components of this relationship by looking specifically at the relationship between TMT absorptive capacity and BI assimilation with the theorization that this relationship will occur through operational-level absorptive capacity. Using the Hoyle and Kenny (1999) approach to test the mediated relationship, we first look at the indirect effect which is strong and significant (0.21, $p < 0.001$) (see Table 4). We subsequently test whether the mediation is partial or full (Baron and Kenny 1986). Results, as shown in Table 5, confirm full mediation. This provides evidence of the link between cultural controls and other core MCS controls provided by BI, but also reflects the importance of line and operational managers achieving this interrelationship and successful integration of MCS as a package.

Hypothesis 3 predicts that IT infrastructure sophistication will enhance BI assimilation. The results shown in Figure 2 and Table 4 support the hypothesis (0.21, $p < 0.05$). This supports the belief that underlying enterprise systems-related IT infrastructure plays an important role in BI assimilation.

Hypothesis 4 predicts increased levels of operational-level absorptive capacity will enhance levels of IT infrastructure sophistication. The results in Figure 2 and Table 4 show a strong and significant relationship (0.40, $p < 0.001$) supporting the theorized relationship. This finding highlights the key role of operational and line managers in driving development of appropriate IT infrastructure.

Hypothesis 5 is the only hypothesis not supported. Hypothesis 5 predicts that increased levels of TMT absorptive capacity will enhance levels of IT infrastructure sophistication. This finding is surprising, given that several earlier MCS studies report a strong relationship between TMT involvement and successful implementation of specific MCS components. We conduct *ad hoc* tests to determine if this relationship is offset by operational-level absorptive capacity, which we found to be strongly linked to IT infrastructure sophistication. The most likely scenario from a theoretical perspective would be that TMT absorptive capacity influences IT infrastructure sophistication indirectly through operational level absorptive capacity. Tests for this mediation effect, using the same techniques for testing H2a, are significant, and reflect full mediation (see Tables 4 and 5).

While this latter finding is inconsistent with our *a priori* theoretical model, the finding is not

TABLE 5
Nested Model Comparison

Direct Paths	R ² in Nested Model (No Direct Path)	R ² in Models with Direct Path	The Magnitude of the Change in R ² (f^2)	Pseudo F = $f^2 \times (n - k - 1)$	Conclusion
TMT's absorptive capacity to BI assimilation: An additional test for (H2a)	0.228	0.231	0.004	1.33	Not significant
TMT's absorptive capacity to IT infrastructure: An additional test for (H5)	0.138	0.139	0.001	0.40	Not significant

necessarily inconsistent with an absorptive-capacity view of assimilation. Prior studies focus on TMT involvement and specific MCS implementation, whereas we focus on absorptive capacity and broad-based MCS implementation. Our findings could be interpreted as indicating the critical importance that operational and line managers play in the integration of technology with business processes and the importance at the user level of recognizing where business value can be derived. This finding should be investigated further in future studies.

V. DISCUSSION AND CONCLUSION

Recent research indicates that MCS are evolving to provide a more strategic view of value-chain activities including product development, sales and marketing, customer relations, and performance management (Davila and Foster 2005, 2007). In just the last few years, we have seen a shift in the fundamental role of the management accountant to consistently include internal analysis and risk management activities (DeLoo et al. 2009). This is reflected in the Institute of Management Accountants' (IMA) December 2008 release of a Statement on Management Accounting that changed the definition of a management accountant for the first time in 30 years (IMA 2008, 1):

Management accounting is a profession that involves partnering in management decision making, devising planning and performance management systems, and providing expertise in financial reporting and control to assist management in the formulation and implementation of an organization's strategy.

In this study, we examine the assimilation of BI systems that automate MCS as a package and support an ever-increasing number of these contemporary responsibilities for management accountants. We examine BI systems provided by one vendor whose product is considered a leader in providing MCS capability. We focus our study on the cultural controls associated with knowledge creation (i.e., absorptive capacity). The resulting organizational absorptive capacity impacts successful assimilation of BI systems.

Our results provide evidence that increased absorptive capacity among operational-level managers is strongly related to both increased levels of BI assimilation and increased levels of sophistication in the underlying technology infrastructure that enables BI systems. At the same time, operational-level absorptive capacity is strongly tied to TMT's absorptive capacity, providing evidence of the importance of cultural controls related to knowledge management and resource development. These results have implications not only for the role of cultural controls in fostering an effective knowledge culture, but also for the view among many researchers that MCS should be considered as a package—i.e., recognizing that specific MCS components are impacted by the surrounding MCS controls as a whole (Chenhall 2003; Malmi and Brown 2008).

The results for BI assimilation provide some evidence of a potential diffusion effect on the management accounting function as the scope of MCS broadens. Several researchers have observed indications of this diffusion effect due to the adoption of enterprise systems and associated increased capability for MCS (see Granlund and Malmi 2002; Caglio 2003; Scapens and Jazayeri 2003; Quattrone and Hopper 2005). Our results indicate that organizations with greater absorptive capacity assimilate BI across business operations, customer relations, and marketing and sales—a phenomenon that could be perceived as showing widespread change consistent with these observations in earlier studies of specific organizations. Future research should explore this diffusion phenomenon in order to better understand the impact on management accountants' roles in organizations and the MCS function overall.

Our research also contributes to understanding the roles of TMT in MCS innovations through an expanded view of organizational absorptive capacity. In particular, we shift the focus of absorptive capacity to a dynamic perspective of knowledge and incorporate a measurement of intensity of effort that has not been considered in prior research on the link between TMT involve-

ment and MCS innovation success. Focusing on this dynamic view of absorptive capacity, we incorporate the knowledge-creation capabilities that are critical to adapting to new information and new technologies. Incorporating this broader measure of absorptive capacity, we are able to provide evidence that both operational-level and TMT absorptive capacity are likely to influence the assimilation of a broad range of strategic systems related to MCS. At the same time, we also show evidence that the absorptive capacities of the operational-level managers are fundamental to establishing the robust IT infrastructure necessary to support strategic systems that support MCS innovations. Operational-level managers' absorptive capacity is similarly key to organizational assimilation of strategic technologies to gain strategic advantage. Prior research has failed to establish the link between TMT's knowledge and assimilation of strategic systems (e.g., Armstrong and Sambamurthy 1999). Our results confirm the TMT effect as indirect and flowing through the operational level. Future research should further explore this indirect effect in order to better understand the mechanisms through which TMT's knowledge can be leveraged through cultural controls. Further, we have focused on one particular aspect of cultural controls, absorptive capacity. Future research should also consider the broader range of cultural controls and the effect on successful assimilation of MCS.

The findings related to absorptive capacity effects on IT infrastructure are also of interest. There is a contrast between operational-level absorptive capacity, which has a significant influence on IT infrastructure sophistication and TMT's absorptive capacity, which does not have a direct effect. *Post hoc* tests reveal an indirect relationship between TMT's absorptive capacity and IT infrastructure with operational-level managers' absorptive capacity fully mediating the relationship. These findings indicate that sophisticated IT infrastructure is developed through socially complex processes that involve collaborations between TMT, IT specialists, and operational-level managers. The cultural controls put in place by TMT have a substantial effect on these collaborations. Moreover, the results also support the calls made in both the accounting and information systems literature for management accountants, line managers, and IT managers to develop broader overlapping knowledge, interpersonal skills, and deeper understanding of the strategic content of the organization (Todd et al. 1995; Nelson and Coopridge 1996; Caglio 2003; Byrd et al. 2004; Dechow and Mouritsen 2005; Quattrone and Hopper 2005; Arnold 2006; Rikhardsson and Kraemmergaard 2006).

In weighing the results of this study, several inherent limitations should be considered. First, although an attempt was made to solicit multiple respondents from each of the targeted firms, only a single contact was available for many of the firms. However, tests show no differences between mean responses received from single respondents versus those from multiple respondents. Second, using the same informants to answer questions on both operational- and strategic-level constructs may result in a respondent's bias or knowledge bias. However, techniques used in the study help to alleviate some of that concern. These include providing respondents with a "No Basis for Answering" option in the survey, capturing the same data from multiple respondents, testing for consistency between multiple respondents, testing for common method variance, and testing for discriminant validity. All of these tests suggest that such concern does not threaten the validity of the results. Future studies could consider using two separate surveys. One would be answered by TMT members and include questions on the TMT characteristics, and the other by middle-/operational-level managers and include questions on BI assimilation, IT infrastructure, etc. Third, the study investigates TMT knowledge and argues that it enables the cultural controls that are necessary for building organizational absorptive capacity, appropriate IT resources, and supporting BI assimilation. Future research could examine the direct effects of cultural controls on the cybernetic and planning controls that are most enabled by BI systems.

This research has significant practical implication for management accountants, organizational management, managers using MCS, IT consultants, and technology vendors. The clearest

message is that successful provision of MCS capability is not achieved by acquiring “state-of-the-art” software, but arises through developing appropriate organizational capability for generating and using new knowledge to support MCS. This reasoning also emphasizes that technology innovation is driven from the bottom-up by operational managers, as opposed to the dominant view of top-down, top-management-driven innovation. The top-down view is traditionally perceived as the driving force behind MCS capability. The findings from this study also reinforce organizations’ continuous effort in knowledge-creation activities at both TMT and operational levels. Organizations should promote knowledge-creation activities as an ongoing process in preference to purposeful and directed short-term knowledge acquisition such as training programs.

REFERENCES

- Anderson, S. W., and S. M. Young. 1999. The impact of contextual and process factors on the evaluation of activity-based costing systems. *Accounting, Organizations and Society* 24: 525–559.
- Armstrong, C. P., and V. Sambamurthy. 1999. Information technology assimilation in firms: The influence of senior leadership and IT infrastructures. *Information Systems Research* 10 (4): 304–327.
- Arnold, V. 2006. Behavioral research opportunities: Understanding the impact of enterprise systems. *International Journal of Accounting Information Systems* 7 (1): 7–17.
- Attewell, P. 1992. Technology diffusion and organizational learning: The case of business computing. *Organization Science* 3 (1): 1–19.
- Baron, R. M., and D. A. Kenny. 1986. The moderator-mediator variable distinction in social psychological-research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 51 (6): 1173–1182.
- Bassellier, G., I. Benbasat, and B. H. Reich. 2003. The influence of business managers’ IT competence on championing IT. *Information Systems Research* 14: 317–336.
- Becerra-Fernandez, I., and R. Sabherwal. 2001. Organizational knowledge management: A contingency perspective. *Journal of Management Information Systems* 18 (1): 23–55.
- Bhimani, A. 2003. A study of the emergence of management accounting system ethos and its influence on perceived system success. *Accounting, Organizations and Society* 28: 523–548.
- Boynton, A. C., R. W. Zmud, and G. C. Jacobs. 1994. The influence of IT management practice on IT use in large organizations. *Management Information Systems Quarterly* 18 (3): 299–318.
- Brignall, S., and J. Ballantine. 2004. Strategic enterprise management systems: New directions for research. *Management Accounting Research* 15: 225–240.
- Broadbent, M., P. Weill, and D. Clair. 1999. The implications of information technology infrastructure for business process redesign. *Management Information Systems Quarterly* 23 (2): 159–182.
- Byrd, T. A., B. R. Lewis, and D. E. Turner. 2004. The impact of IT personnel skills on IS infrastructure and competitive IS. *Information Resources Management Journal* 17 (2): 38–52.
- Byrne, S., and B. Pierce. 2007. Towards a more comprehensive understanding of the roles of management accountants. *European Accounting Review* 16 (3): 469–498.
- Caglio, A. 2003. Enterprise resource planning systems and accountants: Towards hybridization? *European Accounting Review* 12 (1): 123–153.
- Carte, T. A., A. B. Schwarzkopf, T. M. Shaft, and R. W. Zmud. 2005. Advanced business intelligence at Cardinal Health. *MIS Quarterly Executive* 4 (4): 413–424.
- Chapman, C. S. 2005. Not because they are new: Developing the contribution of enterprise resource planning systems to management control research. *Accounting, Organizations and Society* 30: 685–689.
- , and L.-A. Kihn. 2009. Information systems integration, enabling control and performance. *Accounting, Organizations and Society* 34: 151–169.
- Chatterjee, D., R. Grewal, and V. Sambamurthy. 2002. Shaping up for e-commerce: Institutional enablers of the organizational assimilation of Web technologies. *Management Information Systems Quarterly* 26 (2): 65–89.
- Chenhall, R. H. 2003. Management control systems design within its organizational context: Findings from contingency-based research and directions for the future. *Accounting, Organizations and Society* 28:

- 127–168.
- . 2005. Integrative strategic performance measurement systems, strategic alignment of manufacturing, learning and strategic outcomes: An exploratory study. *Accounting, Organizations and Society* 30: 395–422.
- , and K. J. Euske. 2007. The role of management control systems in planned organizational change: An analysis of two organizations. *Accounting, Organizations and Society* 32: 601–637.
- Chin, W. W. 1998. The partial least squares approach to structural equation modeling. In *Research Methods for Business Research*, 22 (1): edited by G. A. Marcoulides. Mahwah, NJ: Lawrence Erlbaum Associates.
- , and P. Newsted. 1999. Structural equation modeling analysis with small samples using partial least squares. In *Statistical Strategies for Small Sample Research*, 307–341, edited by R. Hoyle. Thousand Oaks, CA: Sage.
- , B. L. Marcolin, and P. R. Newsted. 2003. A partial least squares latent variable modeling approach for measuring interaction effects: Results from a Monte Carlo simulation study and an electronic-mail emotion/adoption study. *Management Information Systems Quarterly* 14 (2): 189–217.
- Choi, B., and H. Lee. 2002. Knowledge management strategy and its link to knowledge creation process. *Expert Systems with Applications* 23 (3): 173–187.
- Churchill, G. A. 1979. A paradigm for developing better measures of marketing constructs. *JMR, Journal of Marketing Research* 16 (1): 64–73.
- Chwelos, P., I. Benbasat, and A. Dexter. 2001. Research report: Empirical test of an EDI adoption model. *Information Systems Research* 12 (3): 304–321.
- Cohen, W. M., and D. A. Levinthal. 1990. Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly* 35 (1): 128–152.
- Cook, S. D. N., and J. S. Brown. 1999. Bridging epistemologies: The generative dance between organizational knowledge and organizational knowing. *Organization Science* 10 (4): 381–400.
- Daft, R. L., and K. E. Weick. 1984. Toward a model of organizations as interpretive systems. *Academy of Management Review* 9: 284–295.
- Daghfous, A. 2004. Absorptive capacity and the implementation of knowledge-intensive best practices. *SAM Advanced Management Journal* 69 (2): 21–27.
- Davenport, T., and L. Prusak. 1998. *Working Knowledge: How Organizations Manage What They Know*. Boston, MA: Harvard Business School Press.
- Davila, A., and G. Foster. 2005. Management accounting systems adoption decisions: Evidence and performance implications from early-stage/startup companies. *The Accounting Review* 80 (4): 1039–1068.
- , and ———. 2007. Management control systems in early-stage startup companies. *The Accounting Review* 82 (4): 907–937.
- Dechow, N., and J. Mouritsen. 2005. Enterprise resource planning systems, management control and the quest for integration. *Accounting, Organizations and Society* 30: 691–733.
- DeLoo, I., B. Verstegen, and D. Swagerman. 2009. Towards a better understanding of management accountants' roles in the Netherlands: The management accountants survey 2007. Working paper, Open University of the Netherlands. Available at SSRN: <http://ssrn.com/abstract=1368119>.
- Dillman, D. A. 2000. *Mail and Internet Surveys: The Tailored Design Method*. 2nd edition. New York, NY: John Wiley and Sons, Inc.
- Ditillo, A. 2004. Dealing with uncertainty in knowledge-intensive firms: The role of management control systems as knowledge integration mechanisms. *Accounting, Organizations and Society* 29: 401–421.
- Dowling, C. 2009. Appropriate audit support system use: The influence of auditor, audit team, and firm factors. *The Accounting Review* 84 (3): 771–810.
- Elbashir, M., and S. Williams. 2007. BI impact: The assimilation of business intelligence into core business processes. *Business Intelligence Journal* 12: 45–54.
- Fichman, R. G., and C. F. Kemerer. 1997. The assimilation of software process innovations: An organizational learning perspective. *Management Science* 43 (10): 1345–1363.
- Fornell, C. D., and F. Larcker. 1981. Evaluating structural equation models with unobservable variables and measurement errors. *JMR, Journal of Marketing Research* 18: 39–50.
- Gefen, D., D. W. Straub, and M. Boudreau. 2000. Structural equation modeling and regression: Guidelines

- for research practice. *Communications of The Association for Information Systems* 4 (Article 7): 1–77.
- Gnatovich, R. 2007. Making a case for business analytics. *Strategic Finance* (February): 47–51.
- Granlund, M., and T. Malmi. 2002. Moderate impact of ERPs on management accounting: A lag or permanent outcome? *Management Accounting Research* 13: 299–321.
- Grant, R. M. 1996. Toward a knowledge-based theory of the firm. *Strategic Management Journal* 17: 109–122.
- Hair, J. F., R. E. Anderson, R. L. Tatham, and W. C. Black. 1998. *Multivariate Data Analysis*. Upper Saddle River, NJ: Prentice Hall.
- Hall, M. 2008. The effect of comprehensive performance measurement systems on role clarity, psychological empowerment and managerial performance. *Accounting, Organizations and Society* 33: 141–163.
- Hambrick, D. C. 1981. Strategic awareness within top management teams. *Strategic Management Journal* 2 (3): 263–279.
- . 1995. Fragmentation and the other problems CEOs have with their top management teams. *California Management Review* 37 (3): 110–127.
- Hambrick, D. C., and P. A. Mason. 1984. Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review* 9 (2): 193–206.
- Howard, P. 2003. *Analytics Volume 1: An Evaluation and Comparison*. Edited by B. Smithson. Milton Keynes, U.K.: Bloor Research.
- Hoyle, R. H., and D. A. Kenny. 1999. Sample size, reliability, and tests of statistical mediation. In *Statistical Strategies for Small Sample Research*, edited by R. H. Hoyle. Thousand Oaks, CA: Sage Publications.
- Huber, G. P., and D. J. Power. 1985. Retrospective reports of strategic-level managers: Guidelines for increasing their accuracy. *Strategic Management Journal* 6 (2): 171–180.
- Hulland, J. 1999. Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic Management Journal* 20 (2): 195–204.
- Hurst, D. K., J. C. Rush, and R. E. White. 1989. Top management teams and organizational renewal. *Strategic Management Journal* 10: 87–105.
- Institute of Management Accountants. 2008. *Statement on Management Accounting: Definition of Management Accounting*. Montvale, NJ: IMA.
- Jansen, J. J. P., F. A. Van den Bosch, and H. W. Volberda. 2005. Managing potential and realized absorptive capacity: How do organizational antecedents matter? *Academy of Management Journal* 48 (6): 999–1015.
- Kakabadse, A., J. McMaho, and A. Myers. 1995. Correlates of internal and external leadership of top management teams: An international comparative study. *Leadership and Organization Development Journal* 16 (7): 10–17.
- Kay, M. 2006. The BPM convergence. *Strategic Finance* (September): 51–55.
- Keen, P. G. W. 1991. *Shaping the Future: Business Design Through Information Technology*. Boston, MA: Harvard Business School Press.
- Kim, L. 1998. Crisis construction and organizational learning: Capability building in catching-up at Hyundai Motor. *Organization Science* 9 (4): 506–521.
- Kippenberger, T. 1997. Organizational learning starts at the top. *The Antidote* 2 (7): 19–22.
- Kline, R. B. 2006. Reverse arrow dynamics. Formative measurement and feedback loops. In *Structural Equation Modeling: A Second Course*, edited by G. R. Hancock, and R. O. Mueller, 43–68. Charlotte, NC: Information Age Publishing.
- Liang, H., N. Saraf, Q. Hu, and Y. Xue. 2007. Assimilation of enterprise systems: The effect of institutional pressures and the mediating role of top management. *Management Information Systems Quarterly* 31 (4): 59–88.
- Lu, H.-P., and S. Wang. 2008. The role of internet addiction in online game loyalty: An exploratory study. *Internet Research* 18 (5): 499–519.
- Malmi, T., and D. A. Brown. 2008. Management control systems as a package—Opportunities, challenges and research directions. *Management Accounting Research* 19: 287–300.
- Mathieson, K., E. Peacock, and W. Chin. 2001. Extending the technology acceptance model: The influence of perceived user resources. *The Data Base for Advances in Information Systems* 32 (3): 86–112.
- McCarthy, W. E. 1982. The REA accounting model: A generalized framework for accounting systems in a

- shared data environment. *The Accounting Review* 57 (3): 554–578.
- Nambisan, S., R. Agarwal, and M. Tanniru. 1999. Organizational mechanisms for enhancing user innovation in information technology. *Management Information Systems Quarterly* 23 (3): 365–395.
- Naranjo-Gil, D., and F. Hartmann. 2006. How top management teams use management accounting systems to implement strategy. *Journal of Management Accounting Research* 18: 21–53.
- , and ———. 2007. Management accounting systems, top management team heterogeneity and strategic change. *Accounting, Organizations and Society* 32: 735–756.
- Nelson, K. M., and J. G. Coopridge. 1996. The contribution of shared knowledge to IS group performance. *Management Information Systems Quarterly* 20 (4): 409–432.
- Nonaka, I. 1994. A dynamic theory of organizational knowledge creation. *Organization Science* 5 (1): 14–37.
- , P. Byosiére, C. C. Borucki, and N. Konno. 1994. Organizational knowledge creation theory: A first comprehensive test. *International Business Review* 3 (4): 337–351.
- , and H. Takeuchi. 1995. *The Knowledge-Creating Company*. New York, NY: Oxford University Press.
- , and N. Konno. 1998. The concept of Ba: Building a foundation for knowledge creation. *California Management Review* 40 (3): 1–15.
- , and R. Toyama. 2003. The knowledge-creating theory revisited: Knowledge creation as a synthesizing process. *Knowledge Management Research and Practice* 1: 2–10.
- Pavlou, P. A., and D. Gafen. 2005. Psychological contract violation in online marketplaces: Antecedents, consequences, and moderating roles. *Information Systems Research* 16 (4): 372–399.
- Podsakoff, P. M., and D. W. Organ. 1986. Self-reports in organizational research: Problems and prospects. *Journal of Management* 12 (4): 531–544.
- Podsakoff, P. M., S. B. MacKenzie, and J.-Y. Lee. 2003. Common method bias in behavioral research: A critical review of the literature and recommended remedies. *The Journal of Applied Psychology* 88 (5): 879–903.
- Porter, M. E. 1985. *Competitive Advantage*. New York: Free Press.
- Quattrone, P., and T. Hopper. 2005. A “time-space odyssey:” Management control systems in two multinational organizations. *Accounting, Organizations and Society* 30: 735–764.
- Rikhardsson, P., and P. Kraemmergaard. 2006. Identifying the impacts of enterprise system implementation and use: Examples from Denmark. *International Journal of Accounting Information Systems* 7 (1): 36–49.
- Ravichandran, T. 2005. Organizational assimilation of complex technologies: An empirical study of component-based software development. *IEEE Transactions on Engineering Management* 52 (2): 249–268.
- Robertson, B., A. Boehler, and J. Hansel. 2007. Sustainable performance improvement through predictive technologies. *Strategic Finance* (June): 57–64.
- Rom, A., and C. Rohde. 2007. Management accounting and integrated information systems: A literature review. *International Journal of Accounting Information Systems* 8 (1): 40–68.
- Scapens, R. W., and M. Jazayeri. 2003. ERP systems and management accounting change: Opportunities or impacts? A research note. *European Accounting Review* 12 (1): 201–233.
- Sethi, V., and W. R. King. 1994. Development of measures to assess the extent to which an information technology application provides competitive advantage. *Management Science* 40 (12): 1601–1627.
- Simons, R. 1995. *Levers of Control*. Boston, MA: Harvard University Press.
- Straub, D. W. 1989. Validating instruments in MIS research. *Management Information Systems Quarterly* 13 (2): 147–169.
- Subramani, M. 2004. How do suppliers benefit from information technology use in supply chain relationships? *Management Information Systems Quarterly* 28 (1): 45–73.
- Sutton, S. G. 2000. The changing face of accounting in an IT-driven environment. *International Journal of Accounting Information Systems* 1 (1): 1–8.
- Szulanski, G. 1996. Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic Management Journal* 17: 27–43.
- Todd, P. A., J. D. McKeen, and R. B. Gallupe. 1995. The evolution of IS job skills: A content analysis of IS job advertisements from 1970 to 1990. *Management Information Systems Quarterly* 19 (1): 1–27.

- Ulrich, D., and M. F. Wiersema. 1989. Gaining strategic and organizational capability in a turbulent business environment. *The Academy of Management Executive* 3 (2): 115–122.
- Van den Bosch, F. A. J., H. W. Volberda, and M. de Boer. 1999. Coevolution of firm absorptive capacity and knowledge environment: Organizational forms and combinative capabilities. *Organization Science* 10 (5): 551–568.
- Vandenbosch, B. 1999. An empirical analysis of the association between the use of executive support systems and perceived organizational competitiveness. *Accounting, Organizations and Society* 24 (1): 77–92.
- Venkatraman, N. 1989. The concept of fit in strategy research: Toward verbal and statistical correspondence. *Academy of Management Review* 14 (3): 423–444.
- Vera-Munoz, S. C., J. L. Ho, and C. W. Chow. 2006. Enhancing knowledge sharing in public accounting firms. *Accounting Horizons* 20 (2): 133–156.
- von Hippel, E. 1994. Sticky information and the locus of problem solving: Implications for innovation. *Management Science* 40 (4): 429–439.
- Williams, S. 2008. Power combination: Business intelligence and the balanced scorecard. *Strategic Finance* 89 (May): 1–8.
- . 2004. Delivering strategic business value: Business intelligence can help management accounting reclaim its relevance and rightful role. *Strategic Finance* 86 (2): 41–48.
- , and N. Williams. 2007. *The Profit Impact of Business Intelligence*. London, U.K.: Elsevier Inc.
- Wu, W. Y., C. Y. Chiang, and J. S. Jiang. 2002. Interrelationships between TMT management styles and organizational innovation. *Industrial Management & Data Systems* 102 (3/4): 10–17.
- Zahra, S. A., and G. George. 2002. Absorptive capacity: A review, reconceptualization, and extension. *Academy of Management Review* 27 (2): 185–203.
- Zhu, K., and K. L. Kraemer. 2002. E-commerce metrics for net-enhanced organizations: Assessing the value of e-commerce to firm performance in the manufacturing Sector. *Information Systems Research* 13 (3): 275–295.

Disaggregating Management Forecasts to Reduce Investors' Susceptibility to Earnings Fixation

W. Brooke Elliott

Jessen L. Hobson

Kevin E. Jackson

University of Illinois at Urbana–Champaign

ABSTRACT: This study examines disaggregated management forecasts as a mechanism to reduce investors' fixation on announced earnings. Our experimental results suggest that investors' earnings fixation is reduced when they initially observe a disaggregated management forecast (earnings and its components) versus when they observe an aggregated forecast (earnings only). We also provide theory-consistent evidence that this reduction in earnings fixation is associated with investors interpreting the summary net income figure as one of several similarly important evaluation inputs rather than a substantially more important input (relative to its components). Finally, we provide evidence that suggests our results are not bounded by the level of emphasis on net income in the subsequent earnings announcement, and not fully explained by three plausible alternative explanations. Our study extends the voluntary disclosure literature by providing evidence that the form of management disclosures can influence investors' interpretation of subsequently announced information, and contributes to practice by providing a potential alternative to stopping earnings guidance.

Keywords: *earnings fixation; forecast disaggregation; stopping guidance.*

Data Availability: *Data available for reproduction purposes.*

I. INTRODUCTION

Market observers (McCafferty 2007), policy makers (CFA Centre for Financial Market Integrity 2006; Herz 2006), investor advocate groups (CFA Centre for Financial Market Integrity 2006, 2007, 2008) and accounting researchers (e.g., Graham et al. 2005; Chen et al. forthcoming; Cheng et al. 2007; Houston et al. 2010) have each expressed concerns about the

We thank Wendy Bailey, Rob Bloomfield, Annie Farrell, Susan Krische, Laura Li, Bob Libby, Mark Nelson, Mark Peecher, Jane Thayer, Brad Tuttle, Martin Wu, two anonymous reviewers from the American Accounting Association's 2008 ABO Research Conference, and workshop participants at Cornell University, University of Illinois at Urbana–Champaign, University of South Carolina, two anonymous reviewers, and Jim Hunton (associate editor) for helpful comments. We also thank Eric Gooden, Cathy Gray, Ling Harris, Walied Keshk, and Ryan Tate for valuable research assistance, and the financial assistance of the Department of Accounting at Florida State University and the University of Illinois Campus Research Board.

Editor's note: Accepted by James Hunton.

Submitted: May 2009

Accepted: June 2010

Published Online: January 2011

tendency of investors to focus on companies' short-term results in lieu of long-term potential. Often this short-termism is attributed to investors' excessive reliance on accounting earnings as the primary input to investment judgments, without fully considering other information that may be relevant to evaluating the company's investment prospects (Graham et al. 2005; CFA Centre for Financial Market Integrity 2006). Further, this excessive reliance on accounting earnings (i.e., earnings fixation) is often cited as the key source of managers' myopic behavior and associated reduction in long-term value delivered to shareholders (CFA Centre for Financial Market Integrity 2006, 2007).¹

Although earnings fixation represents an institutional phenomenon that can be undesirable to managers and investors, prior research has primarily focused on documenting the existence and extent of earnings fixation rather than on mechanisms that reduce this fixation. We seek to fill this gap by examining a mechanism to reduce earnings fixation. Specifically, we examine whether and how the level of aggregation of a management forecast can reduce the extent to which investors' investment judgments reflect earnings fixation.

We focus on management forecasts because (1) they are often investors' first exposure to a firm's expectation of its future bottom-line summary earnings, and (2) management has full discretion over the format of the forecast—specifically, the items for which a forecast is provided. Furthermore, several stakeholder groups have advocated the potentially costly approach of eliminating management forecasts to reduce fixation (Chen et al. forthcoming; Houston et al. 2010). We examine a potential alternative to eliminating guidance as a mechanism to reduce earnings fixation: simply altering the level of aggregation in the guidance provided.

We predict that investors' susceptibility to earnings fixation will decrease (increase) when investors initially observe a management forecast that disaggregates (aggregates) the components of forecasted earnings. Our prediction is guided by research in psychology and marketing that suggests that because information can often be perceived in several different ways, the interpretation of that information depends on the particular concept or knowledge structure that is currently active (Higgins and King 1981; Wyer and Srull 1981; Yi 1990a, 1990b).

More specifically, we contend that investors possess at least two general concepts related to net income—one conceptualization of net income as one of several important inputs to evaluating a company's prospects, and another as a sufficient and convenient summary measure, and thus a substantially more important measure, of a company's prospects (relative to its components). We predict that when investors observe a disaggregated earnings forecast, which places net income among a list of other potentially relevant inputs, the former conceptualization will likely be primed. In contrast, when investors observe an aggregated earnings forecast, where net income is isolated, the latter conceptualization of net income will likely be primed. Once the investor observes net income in the subsequent earnings announcement, the primed knowledge structure will be activated and used to evaluate net income. Thus, the investment judgments of investors who initially observe a *disaggregated* (an *aggregated*) earnings forecast are expected to display less (more) earnings fixation because they evaluate net income as one of several similarly important inputs (as a substantially more important input) in assessing the company's prospects.

We use an experiment to test these predictions. Participant-investors start by observing a management earnings forecast and making initial investment judgments. We manipulate the forecast such that it is *disaggregated* or *aggregated*. Consistent with Hirst et al. (2007), disaggregated earnings forecasts are defined as forecasts of the bottom-line net income number as well as

¹ The CFA Centre for Financial Market Integrity's (2006) report titled "Breaking the Short-Term Cycle" defines short-termism as the "excessive focus of some corporate leaders, investors, and analysts on short-term, quarterly earnings and a lack of attention to the strategy, fundamentals, and conventional approaches to long-term value creation."

forecasts of each line item on the income statement (e.g., revenues, research and development expenses, etc.). Aggregated earnings forecasts are defined as forecasts of the bottom-line net income number alone. Participants then observe an earnings announcement and make a second set of investment judgments. Although the earnings announcement always reports a net income figure that matches forecasted net income, we manipulate whether the underlying components of net income indicate *more favorable* or *less favorable* performance. This research design allows us to infer earnings fixation when participants' investment judgments are relatively the same regardless of the favorability indicated by the net income components, as such, judgments merely reflect the fact that the announced net income figure is identical across favorability conditions. In contrast, fixation is taken to be reduced when participants' judgments are more (less) favorable when the reported net income components indicate more (less) favorable performance.

As predicted, participants' investment judgments reflect less earnings fixation when they initially observe a disaggregated management earnings forecast. Specifically, participants who initially observe a disaggregated forecast provide investment judgments that are more (less) favorable when the net income components are more (less) favorable. In contrast, participants who initially observe an aggregated forecast provide investment judgments that are insensitive to the reported levels of the net income components. This evidence suggests that the level of aggregation in the initial management earnings forecast influences the extent to which participants' investment judgments reflect line items other than bottom-line net income (i.e., the extent of earnings fixation).

We also find evidence consistent with our theory that this reduction in earnings fixation is associated with investors' interpreting the summary net income figure as one of several similarly relevant evaluation inputs rather than a substantially more important input (relative to its components). We ask participants to allocate 100 points among the bottom-line net income number and its components to indicate each item's importance in evaluating the company as a potential investment. Participants who observe a disaggregated forecast assign less (more) importance to the bottom-line net income number (net income components) than those who observe an aggregated forecast. This result supports our contention that the disaggregated forecast leads investors to consider net income as one of several similarly important inputs in evaluating the company as a potential investment.

In additional analyses, we find evidence that that our results cannot be fully explained by (1) differences in available information across forecast aggregation levels, (2) management signaling precisely which financial statement items are relevant by disaggregating the forecast, or (3) forecast aggregation levels affecting investors' perceived credibility of the forecast. We also find that the effect of a disaggregated forecast is nearly identical regardless of whether the subsequent earnings announcement emphasizes or deemphasizes the bottom-line net income number. Thus, the effect of disaggregating the earnings forecast does not appear to be bounded by the level of emphasis of the bottom-line net income number in the subsequent earnings announcement.

This study makes several contributions to the extant literature. First, we extend accounting research that documents the existence of investors' fixation on earnings (e.g., Bushee 1998) by proposing a potentially attractive approach to reduce fixation in investors' judgments. Second, we extend the disclosure literature in accounting by providing evidence that benchmarks, or forecasts, cannot only be used to evaluate current and prior period information (Schrand and Walther 2000; Krische 2005), but can also affect how investors interpret subsequently reported accounting information. Specifically, we show that a disaggregated earnings forecast can lead investors to interpret subsequently announced net income to be one of several similarly important inputs in evaluating a company's future prospects. Third, we extend the disclosure literature on the effects of forecast characteristics on investors' responses to subsequently announced earnings (Libby and Tan 1999;

Tan et al. 2002; Libby et al. 2006; Han and Tan 2007) by examining the influence of a different characteristic—the level of forecast aggregation—on investors' response to subsequently announced earnings.

This study also extends recent research on the influence of the level of forecast aggregation on investor judgments. Hirst et al. (2007), for example, report results suggesting that investors judge disaggregated forecasts as more credible than aggregated forecasts and deem subsequently reported net income to be of higher quality. One implication is that investors should rely on subsequently reported earnings more when they initially observe a disaggregated forecast. By showing that the level of aggregation influences not only investors' judgments of forecast credibility, but also the extent to which investors' judgments reflect earnings fixation, we refine our understanding of the relationship between forecast aggregation level and investors' reactions to subsequently reported earnings.

Finally, this study contributes to practice. In particular, we provide evidence that modifying the format of management earnings guidance may reduce investors' susceptibility to earnings fixation, while avoiding the potentially costly effects associated with eliminating guidance. This evidence should be important to both managers, who have incentives to reduce pressures for myopic behavior, and regulators, who have expressed concerns about the effects of investors' earnings fixation.

Section II discusses relevant background information and develops our hypothesis. Section III describes our experiment. Section IV presents our main results along with evidence of our theory and describes an alternative theoretical explanation. Section V presents supplemental analyses. Section VI concludes.

II. BACKGROUND AND HYPOTHESIS DEVELOPMENT

Company managers, financial market participants, regulators, and other stakeholders have each raised concerns about the tendency of investors to fixate on earnings (see Bushee 2001; Graham et al. 2005; CFA Centre for Financial Market Integrity 2007). Earnings fixation refers to investors' excessive reliance on a company's accounting earnings in evaluating the firm, without fully considering other information that is relevant to evaluating the company's investment prospects (see Graham et al. 2005).² Although its effect on the U.S. economy is not easily quantified, investors' earnings fixation and related short-termism is commonly depicted as detrimental to the judgments and decisions of both managers and investors (Bushee 1998, 2001; Graham et al. 2005).³

We argue that earnings fixation occurs when investors conceptualize net income as a substantially more important input (relative to its components) in evaluating a company's prospects, rather than as one of several similarly important inputs. Research in psychology suggests that investors' initial conceptualization of a net income figure is likely to persist when subsequently exposed to the net income figure (Fischhoff 1977). Although announced earnings are undoubtedly important to investors, earnings forecasts often represent investors' initial exposure to a net income amount (*The Economist* 2006; Houston et al. 2010). Thus, our study focuses on earnings forecasts as an opportunity to affect investors' initial conceptualization of net income.

Research in accounting identifies several reasons for companies to issue forecasts. For example, management forecasts help to calibrate analysts' forecasts to minimize analyst forecast errors and forecast dispersion (e.g., Ajinkya and Gift 1984; Houston et al. 2010) and guide

² Our definition of earnings fixation is consistent with the idea of fixating on earnings without regard to how bottom-line net income is derived (Murray 1991).

³ See Graham et al. (2005) for a discussion of managers' myopic decision-making in response to a perceived investor fixation on earnings.

analysts' forecasts toward beatable earnings targets (e.g., Cotter et al. 2006). Management guidance can also reduce litigation risk associated with not warning investors of impending bad news (e.g., Skinner 1994, 1997; Johnson et al. 2001). In addition, issuing forecasts can aid in building a reputation for transparent reporting (e.g., Graham et al. 2005; Hutton and Stocken 2009).

Recently, several stakeholders have advocated the elimination of earnings forecasts, including the CFA Institute and the Business Roundtable, the National Investor Relations Institute, the Aspen Institute's Corporate Values Strategy Group, the Committee for Economic Development, the U.S. Chamber of Commerce, members of Congress, and members of the Securities and Exchange Commission (Janjigian and Ozanian 2006; Houston et al. 2010). Also, several companies have opted to eliminate earnings guidance (e.g., McDonald's Corp. and Merrill Lynch; see also Chen et al. forthcoming; Cheng et al. 2007; Houston et al. 2010). These parties argue that earnings forecasts lead investors to fixate on earnings that, in turn, direct managers to engage in myopic behavior that often runs counter to maximizing the company's long-term growth and shareholder value (Graham et al. 2005; Houston et al. 2010).

Eliminating earnings forecasts could be costly to the company, however. For example, eliminating forecasts may not be a realistic option for smaller companies that are less closely followed by investors (McCafferty 2007). In addition, research in this area finds a negative market reaction and a deterioration of the information environment for firms that stop providing quarterly guidance (Chen et al. forthcoming; Houston et al. 2010). Specifically, Houston et al. (2010) provide evidence of a decrease in analyst coverage of "guidance stoppers," and Bowen et al. (2008) suggest that a decrease in analyst coverage leads to an increase in the cost of raising equity capital.

We propose that, rather than eliminate forecasts, managers might consider changing the type of forecasts they issue.⁴ Managers provide forecasts with different levels of aggregation. For example, a recent examination of S&P 500 companies indicates that of the 63 percent of firms that give earnings guidance, 23 percent provide highly aggregated forecasts that forecast earnings alone, while 22 percent provide highly disaggregated forecasts that forecast several components of earnings (see Lansford et al. [2010] and forecast examples at Exhibit 1). This research suggests that managers' choice of whether to issue aggregated or disaggregated forecasts is a function of their desire to signal their ability to forecast, enhance the credibility of earnings forecasts, and respond to external demand for additional information. We argue, however, that the level of managements' forecast aggregation can also affect the extent of investors' earnings fixation.

We predict that a management forecast that is more disaggregated will reduce earnings fixation. Prior research in psychology and marketing suggests that because information can often be perceived in several different ways, the interpretation of that information depends on the particular concept or knowledge structure that is currently active (Higgins and King 1981; Wyer and Srull 1981; Yi 1990a, 1990b). More specifically, because individuals do not undertake an exhaustive memory search for all potential useful concepts, the active or accessible concepts are more likely to be used to process subsequent information (Wyer and Srull 1981; see also Tversky and Kahneman 1973, 1974; Taylor and Fiske 1979).

More closely related to our study, Yi (1990a, 1990b) examined the way an advertising context

⁴ We acknowledge that there can also be costs to issuing a disaggregated forecast. For example, providing detailed guidance of a firm's operations may assist competitors. In addition, for firms with complex operations, it may be more difficult for its managers to predict the firm's earnings components (see Lansford et al. 2010 for a discussion). Thus, any given manager must weigh the costs associated with stopping guidance against the costs (benefits) of issuing disaggregated guidance.

EXHIBIT 1

Examples of Management Forecasts

Panel A: Aggregated Management Forecast^a

XEROX

- Xerox expects first-quarter 2006 earnings in the range of 20-23 cents per share. The company also reiterated its full-year 2006 guidance of \$1.00-\$1.07 per share. Mulcahy indicated that she now expects the company will deliver full-year earnings in the high end of this range.

Panel B: Disaggregated Management Forecast^b**2006 Outlook**

- **Revenue:** Expected to be between \$9.1 billion and \$9.7 billion.
- **Gross margin:** 59 percent, plus or minus a couple of points (excluding share-based compensation effects of approximately 1 percent).
- **Expenses (R&D plus MG&A):** Approximately \$3.3 billion (approximately \$3 billion excluding share-based compensation effects of approximately \$300 million).
- **Gains from equity investments** and interest and other: Approximately \$140 million.
- **Tax rate:** Approximately 32 percent.
- **Depreciation:** \$1.1 billion, plus or minus \$100 million.
- **Amortization** of acquisition-related intangibles and costs: Approximately \$20 million.

^a Xerox released an aggregated management forecast, which provided a forecast of earnings only.

^b Intel released a disaggregated management forecast, which provided a forecast of earnings components.

could affect the processing of product information in print ads.⁵ Yi (1990a, 1990b) finds that prior exposure to contextual factors can prime certain product attributes and, in turn, increase the likelihood that consumers interpret product information in terms of these activated attributes. Specifically, Yi (1990a) conducts an experiment in which consumers were given an automobile ad that highlights the car's "large size." Noting that consumers likely have knowledge structures of a car's "large size" that depict the attribute either as a favorable input for car safety or an unfavorable input for fuel efficiency, Yi (1990a) argues that the context surrounding the ad can affect which concept or knowledge structure is activated and used. Specifically, the author predicts that placing the ad within an article about safety primes the "car safety" knowledge structure, making it more likely that consumers interpret the car's large size as a favorable input for car safety. Yi's (1990a) results support his prediction. Consumers provided more favorable judgments about the advertised car when it was located within the car safety article than when it was not.

We contend that investors have at least two general concepts or knowledge structures related to net income. Prior research in accounting suggests that investors sometimes conceptualize net income as one of several important inputs to evaluating a company's prospects (see Elliott et al. 2008), likely obtained from formal education (e.g., see Penman 2007), investing guides, or experience. However, many have suggested that investors often forgo this net income conceptualization and instead conceptualize net income as a sufficient and convenient summary measure of a company's prospects (e.g., see CFA Centre for Financial Market Integrity 2007).⁶ The latter conceptualization leads investors to fixate on a bottom-line net income number in subsequent earnings announcements.

We argue that an aggregated forecast isolates net income, priming the conceptualization of net income as a convenient summary measure and, thus, as a substantially more important measure of company prospects relative to its components. By contrast, a disaggregated forecast places net income among a list of other potential inputs in judging the company's prospects. In this case, similar to consumers in Yi's (1990a) study, we propose that a disaggregated forecast places net income in a context that primes or makes salient investors' knowledge structures of net income as one of several similarly relevant inputs in evaluating the company's prospects. Once the investor observes net income in the subsequent earnings announcement, the primed knowledge structure of net income as one of several similarly relevant inputs is activated.⁷ Thus, investors' evaluation of net income in assessing the company's prospects is made in terms of the activated knowledge structure. As such, we expect the investment judgments of investors who initially observe a *disaggregated* (an *aggregated*) earnings forecast to display less (more) earnings fixation because they evaluate net income as one of several similarly relevant inputs (as a substantially more

⁵ Yi (1990a) contemplates both cognitive and affective priming effects on attribute interpretation. Our interest in the effect of forecast display suggests a cognitive priming effect would be more applicable to our study. While Yi (1990a) documents that both cognitive and affective primes can influence attribute interpretation, the author finds that cognitive primes influence attribute interpretation independent of affective primes and that the two types of priming do not interact. Thus, considering affective priming is beyond the scope of our study.

⁶ We note that net income can sometimes be a sufficient measure of a company's prospects. Thus, treating net income as a summary measure of a company's prospects may also result from experiencing the past sufficiency of using net income in this manner.

⁷ While disclosures other than forecasts may also precede any given earnings announcement, or may occur between a forecast and the related earnings announcement, few, if any, of these other disclosures are likely to cue investors to think of the summary net income figure as one of several similarly important inputs as opposed to the most important input (or at least, not to the same extent as forecasts). In addition, forecasts other than management forecasts, such as analyst forecasts, may also cue investors to conceptualize the summary net income figure differently and could either exacerbate or attenuate our predicted effects depending on the level of aggregation of the analyst forecast, the timing of the analyst forecast disclosure relative to the management disclosure, and investors' desire and ability to access the analyst forecast. We focus on management as opposed to analyst forecasts because we are interested in examining mechanisms that managers can use to reduce investor earnings fixation (other than simply eliminating forecasts altogether).

relevant input) in assessing the company's prospects. Thus, we derive our primary hypothesis as follows:

Hypothesis: After receiving a company's earnings announcement, the level of earnings fixation reflected in investors' investment judgments will be lower (higher) when they initially observe a disaggregated (aggregated) earnings forecast.

III. EXPERIMENTAL METHOD

Participants

Two hundred and one students enrolled in graduate business classes from two large state universities participated in the experiment as reasonably informed investors. Our participants have a reasonable understanding of business, accounting, and finance as, on average, participants had taken 7.61 accounting courses and 3.53 finance courses, and 91 percent of participants had used financial statements to evaluate a company's performance at least one time. In addition, 39 percent of participants stated that they had purchased common stock or debt securities, while 88 percent said that they planned to do so in the next five years.

We use students enrolled in graduate business classes as proxies for reasonably informed investors for two primary reasons. First, as Libby et al. (2002, 802) note, sophisticated participants are in short supply and should only be used when the research question necessitates it. We matched participants with the goals of our experiment. Prior to conducting our experiment, we determined that (1) our task was similar to those described in Elliott et al. (2007) as low in integrative complexity, and (2) a reasonable understanding of financial accounting concepts and basic finance would be necessary and sufficient for participants to respond meaningfully to our experimental materials. Second, prior research that uses tasks with similar characteristics to the tasks in this study (and similar to those classified as low in integrative complexity by Elliott et al. [2007]), finds no substantive differences in the responses of professional analysts and graduate business students (e.g., Barton and Mercer 2005).⁸

Experimental Design and Procedures

To investigate our hypothesis, we use a 2×2 between-subjects design. Participants were asked to take on the role of an investor evaluating a hypothetical firm (LearningWare). Participants first received background information that described the company as a developer and seller of various software applications. After examining this background information, participants observed the company's forecast for the year ending December 31, 2007. The forecast was issued on March 30, 2007 and included summary income statements for the years ending December 31, 2005 and 2006 along with the forecast, and suggested a trend in performance consistent with the two previous years. We manipulated the format of the earnings forecast on two levels. Half of the participants received an aggregated earnings forecast (*Aggregated* condition) that aggregated the components of earnings such that only a bottom-line net income measure was provided in the forecast. The remaining participants received a disaggregated forecast (*Disaggregated* condition), in which a forecast was provided for earnings as well as its components (see Appendix A).⁹ After

⁸ In addition, individual investors are an important investor group in their own right. In early 2008, approximately 52.2 million households (45 percent) owned shares of publicly traded stocks, and 16.2 million of these households held stock outside of employer-sponsored plans (Investment Company Institute and the Securities Industry and Financial Markets Association [ICI/SIFMA] 2008). Thus, financial executives and regulators should (and clearly do—<http://sec.gov/about/whatwedo.shtml>) consider this important group of shareholders, as well as professional investors and analysts, in designing and regulating financial disclosures.

⁹ The actual format of the *Aggregated* and *Disaggregated* conditions was adapted from Hirst et al. (2007). Forecasted earnings and its percentage increase were described in text at the top of the forecast press release in the *Aggregated*

examining management's forecast, participants used an 11-point Likert-type scale to indicate the likelihood that they would consider the company as a potential investment and the company's attractiveness as a potential investment (i.e., initial investment judgments are used as covariates in statistical tests).

Upon completing these judgments, participants provided demographic information before continuing the experiment. Participants then received the company's earnings announcement in addition to the same background information and forecast release they had previously seen.¹⁰ The earnings announcements for each condition showed that reported earnings were the same as forecasted earnings.¹¹ However, the summary income statement showed either higher revenue and research and development expense (*Favorable* condition) or lower revenue and research and development expense (*Unfavorable* condition) than would be expected given the trend of the previous two years (Appendix B).¹²

After examining the company's earnings announcement, participants again indicated the likelihood that they would consider the company as a potential investment and the company's attractiveness as a potential investment (i.e., final investment judgments and primary dependent measure). In addition, participants answered several questions that served as manipulation checks and several questions designed to detect evidence of the process they used in forming their judgments.

Highlights of Design Features

To test a positive (versus negative or null) hypothesis for reducing earnings fixation, we examine when participants discern between two companies whose earnings announcements report identical net income, but whose net income components indicate either more or less favorable future prospects. Manipulating revenue to be less (more) than expected is a clear indicator of unfavorable (favorable) future prospects. However, in order to provide a net income figure that does not change, the *Unfavorable* (*Favorable*) setting requires an income-increasing (-decreasing) event that also can be interpreted as unfavorable (favorable). Therefore, we also choose to manipulate R&D expense; although decreasing (increasing) R&D expense increases (decreases) net income, it can be interpreted as an unfavorable (favorable) indicator.

To reduce the likelihood that differences in information across forecast aggregation levels

condition. Forecasts of earnings components (i.e., revenue, gross profit, research and development expense, and earnings) and their percentage changes were described in text at the top of the forecast press release in the *Disaggregated* condition. In addition, as was the case in Hirst et al. (2007), the *Disaggregated* condition included a forecasted summary income statement shown alongside the summary income statements for the prior two years.

¹⁰ Asking participants to provide demographic information between receiving the initial forecast and receiving the company's earnings announcement was done to, at least partially, simulate the natural time delay between when a company issues an initial forecast and the subsequent earnings announcement. We acknowledge that, in the natural setting, the time delay would be longer and that this could potentially weaken the results we report below. However, even in the natural setting, investors have access to previously issued forecasts and are often prompted (by management or some other financial intermediary) to reconsider these forecasts in evaluating the realizations reported in the earnings announcement. Prior accounting studies also have allowed participants to access previously issued forecasts (see Libby et al. 2006; Han and Tan 2007).

¹¹ An alternative design choice would have been to manipulate reported earnings and observe whether participants' judgments reflected the difference in reported earnings. In this case, statistically significant judgment differences would indicate earnings fixation and a null result would indicate reduced earnings fixation. Our choice to have identical earnings for all participants but manipulate the underlying components means that evidence for our hypothesis (i.e., reduced earnings fixation) is associated with a significant statistical finding rather than a null finding. Thus, we believe our design choice provides stronger support for our hypothesis.

¹² In order to examine boundary conditions for the impact of disaggregating the earnings forecast, we also manipulated the format of the earnings announcement on two levels. Half of the participants received an earnings announcement that emphasized net income by highlighting earnings and its percentage change in the statement heading, while the remaining participants received an earnings announcement that deemphasized net income by highlighting the components of earnings along with earnings itself. We describe this boundary condition test in our "Supplemental Analyses" section.

explain our expected findings, our setting is designed to allow participants across *Forecast Aggregation* conditions to develop the same expectations for the 2007 announced income statement. All participants observe previously announced results in a 2007 forecast that indicate net income and its components for 2006 were roughly 20 percent higher than in 2005. In addition, all participants view a forecast of 2007 net income that exceeds 2006 net income by 20 percent. Thus, the expected increase in net income components in 2007 would likely be 20 percent. Our disaggregated forecast merely makes explicit the likely expectations for the 2007 net income components by forecasting values for 2007 net income components that are 20 percent higher than 2006 amounts. In our discussion of alternative explanations in the "Results" section, we provide evidence that participants had similar expectations across *Forecast Aggregation* conditions.

To reduce the likelihood that the impact of disaggregating the 2007 forecast is explained by management signaling which items are most relevant, our disaggregated forecast provides a forecast for all components of net income. Because management in our setting displays no selectivity over which net income components they forecast, the likelihood that participants perceive the disaggregation as a signal of items that are most important is reduced. In addition to the above design choices, we also collect additional data to address the likelihood of these alternative explanations for our findings in our supplemental analyses discussed at the end of the next section.

IV. RESULTS

Manipulation Checks

To determine whether participants perceived different levels of aggregation across the *Aggregated* and *Disaggregated* conditions, we asked participants to assess the level of detail in the forecast they observed. Using an 11-point Likert-type scale anchored on 1 = Not at All Detailed and 11 = Very Detailed, participants gave mean judgments of forecast detail of 3.97 and 6.07 for the *Aggregated* and *Disaggregated* forecast conditions, respectively. As expected, participants judged the aggregated forecast to have less detail than the disaggregated forecast ($t = 6.78, p < 0.01$). In addition, we asked participants to indicate the type of forecast, if any, they observed. Seventy-two percent of participants in the *Aggregated* condition correctly identified that they had received a forecast of net income alone, while 89 percent of participants in the *Disaggregated* condition correctly identified that they had received a forecast of each of the items on the income statement, including net income.¹³ Together, these results suggest that participants were sensitive to our manipulation of *Forecast Aggregation* level.

Initial Investment Judgments

Recall that participants provided initial investment judgments after viewing the management forecast (but before viewing the earnings announcement). Even though participants are exposed to one of our manipulated variables before providing their initial investment judgments, we did not expect the level of aggregation in the management forecast to influence participants' initial investment judgments in a predictable manner. Consistent with our expectation, results (not tabulated) reveal that participants' initial investment judgments are not influenced by the level of aggregation in the management forecast to which they were exposed ($p\text{-value} > 0.50$). Thus, in the analyses reported below we simply include participants' initial investment judgments as a covariate.

The Effects of Forecast Disaggregation

Our hypothesis predicts that investors who initially observe a disaggregated management forecast will exhibit less earnings fixation than investors who initially observe an aggregated

¹³ Our results are inferentially identical if we remove those participants who incorrectly responded to this manipulation check question.

management forecast. We test our hypothesis using participants' judgments of (1) the likelihood of considering the company as a potential investment, and (2) the company's attractiveness as a potential investment as indicated after participants have observed both the forecast and the earnings announcement. A Cronbach coefficient alpha score of 0.92 reveals that the "likelihood of investment" and "investment attractiveness" measures likely measure the same underlying construct. Therefore, we report the results of the hypothesis tests for a single combined investment judgment dependent measure based on the average of the two measures for each participant, adjusted for participants' initial investment judgments (after exposure to only the forecast) as a covariate.¹⁴

Recall that participants in all conditions observe the same forecasted earnings and the same announced earnings that meet the company's forecast. However, the components that underlie earnings differ. Earnings fixation is thus inferred when participants indicate the same investment judgments when the earnings components differ (i.e., across the *Unfavorable* and *Favorable* conditions). Conversely, we infer a reduction in earnings fixation when participants' investment judgments differ as the earnings components differ.

Results of our experiment support the hypothesis. Table 1, Panel A provides adjusted least square means and standard errors for participants' final investment judgments adjusted for their initial investment judgments as a covariate. Using an 11-point Likert-type scale, anchored on 1 = Not at All Likely/Attractive and 11 = Very Likely/Attractive, participants' final investment judgments overall are more favorable in the *Favorable* condition (mean = 7.30) than in the *Unfavorable* condition (mean = 6.69). An overall ANCOVA (see Panel B of Table 1 and Figure 1), using participants' initial investment judgments as a covariate, reveals a significant interaction between the *Unfavorable/Favorable* and the *Aggregated/Disaggregated* conditions ($F = 3.93$, $p = 0.05$, two-tailed).

Table 1, Panel A indicates the adjusted least square mean investment judgments for participants in the *Aggregated* forecast condition are 7.58 and 7.38 for the *Favorable* and *Unfavorable* conditions, respectively. A simple effects test (see Panel C of Table 1) indicates that the investment judgments of participants in the *Aggregated* forecast condition are no different in the *Favorable* condition (where the components of earnings were more favorable) than in the *Unfavorable* condition (where the components of earnings were less favorable; $t = 0.70$, $p = 0.24$, one-tailed). Turning to participants in the *Disaggregated* forecast condition, the adjusted least square mean investment judgments are 7.02 and 6.00 for the *Favorable* and *Unfavorable* conditions, respectively. In this case, a simple effects test (Table 1, Panel C) indicates a significant difference across the *Favorable* and *Unfavorable* conditions ($t = 3.53$, $p < 0.01$, one-tailed).¹⁵ As predicted, participants in the *Disaggregated* forecast condition distinguish between the company's investment potential when its earnings components were more favorable versus less favorable, suggesting that for these participants earnings fixation is reduced.

¹⁴ When analyzing participants' likelihood and attractiveness measures separately, each measure yields the same inferences as the combined investment judgment measure.

¹⁵ *A priori*, the precise pattern of expected cell means is unclear. Our hypothesis specifies that the slope between *Unfavorable* and *Favorable* will differ by the level of aggregation in the management forecast, but it does not suggest a specific pattern of cell means. We expected participants in the *Aggregated/Unfavorable* and *Aggregated/Favorable* conditions to rate investment potential as more favorable than participants in the *Disaggregated/Unfavorable* condition. Specifically, participants in the two former conditions are expected to be more influenced by the realization of the predicted increase in net income, while those in the latter conditions are expected to be more influenced by the failed realizations of predicted increases in revenues and R&D. It is less clear how the investment judgments of participants in the *Disaggregated/Favorable* condition would fall relative to the other conditions. While these participants observed that the company surpassed expectations in reported revenues and spending on R&D, it is somewhat ambiguous as to whether the increased spending on R&D is positive or negative news.

TABLE 1
Descriptive Statistics and Hypothesis Tests

Panel A: Investment Judgment Adjusted Least Squares Means [standard error]^a

	Unfavorable	Favorable	Total
Aggregated	7.38 [0.21] n = 49	7.58 [0.20] n = 50	7.48 [0.15] n = 99
Disaggregated	6.00 [0.20] n = 52	7.02 [0.21] n = 50	6.51 [0.14] n = 102
Column Mean	6.69 [0.14] n = 101	7.30 [0.15] n = 100	6.99 [0.15] n = 201

Panel B: Overall ANCOVA

Source	df	Mean Square	F-statistic	p-value ^b
Unfavorable/Favorable	1	18.35	8.80	<0.01
Forecast Aggregation	1	46.97	22.51	<0.01
Unfavorable/Favorable × Forecast Aggregation	1	8.19	3.93	0.05
Initial Judgments (COV)	1	335.15	160.61	<0.01

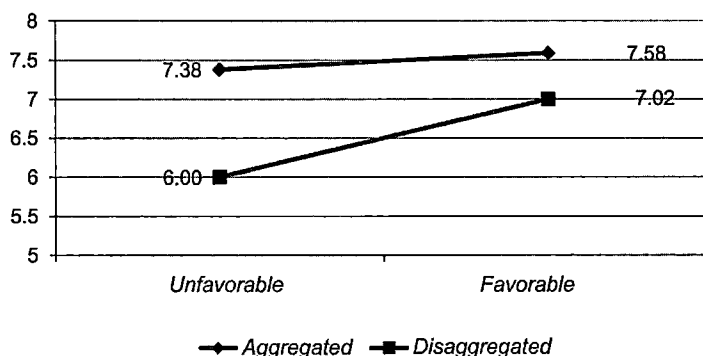
Panel C: Simple Effects Test for Effect of Forecast Aggregation

	df	Mean Square	t-statistic	p-value ^b
Unfavorable is not significantly different from Favorable when the forecast is aggregated	1	1.02	0.70	0.24
Unfavorable is significantly different from Favorable when the forecast is disaggregated	1	25.97	3.53	<0.01

^a Participants judged their likelihood of considering the company as a potential investment (the attractiveness of the company as an investment) using an 11-point scale anchored on 1 = Not at All Likely (Attractive) and 11 = Very Likely (Attractive). Participants made initial judgments after receiving an Aggregated (Disaggregated) management forecast of earnings only (earnings and earnings components) and made the same judgments again after receiving an earnings announcement revealing the realization. Participants observed announced earnings that matched forecasted earnings, but whose revenues and R&D expense depicted either Unfavorable or Favorable performance. The adjusted least square means reported in Panel A were derived by averaging the likelihood and attractiveness responses provided after observing both the forecast and the realization for each participant, adjusted for participants' initial investment judgments (after observing only the forecast) as a covariate.

^b p-values for the overall ANCOVA are two-tailed; p-values for the simple effects tests are one-tailed given a directional prediction.

FIGURE 1
Effects of Forecast Disaggregation on Investment Judgments



Participants judged their likelihood of considering the company as a potential investment (the attractiveness of the company as an investment) using an 11-point scale anchored on 1 = Not at All Likely (Attractive) and 11 = Very Likely (Attractive). Participants made initial judgments after receiving an *Aggregated* (*Disaggregated*) management forecast of earnings only (earnings and earnings components) and made the same judgments again after receiving an earnings announcement revealing the realization. Participants observed announced earnings that matched forecasted earnings, but whose revenues and R&D expense depicted either *Unfavorable* or *Favorable* performance. The adjusted least square means reported above were derived by averaging the likelihood and attractiveness responses provided after observing both the forecast and the realization for each participant, adjusted for participants' initial investment judgments (after observing only the forecast) as a covariate.

Evidence for Our Theory

Our theory suggests that a disaggregated forecast's ability to reduce earnings fixation is associated with investors who observe a disaggregated forecast interpreting net income to be one of several similarly important inputs in evaluating a company's prospects more than investors who observe an aggregated forecast. To examine this theory, we asked participants to allocate 100 points among the different components of the summary income statement, based on how important they deemed each item for evaluating the company as a potential investment. Fewer (more) points allocated to net income (net income components) would be consistent with participants viewing net income as one of several evaluation inputs rather than as a sufficient summary measure of the company's prospects.

The mean points allocated to net income are 26.30 and 22.82 for participants in the *Aggregated* and *Disaggregated* forecast conditions, respectively (see Table 2, Panel A). Results of an ANOVA test (Table 2, Panel B) reveal a main effect for forecast aggregation ($F = 3.44$, $p = 0.03$, one-tailed), indicating that participants in the *Disaggregated* condition allocate fewer points to net income than participants in the *Aggregated* condition. The result supports our theory about why disaggregated forecasts reduce earnings fixation.¹⁶

¹⁶ Although an examination of the means reported in Panel A of Table 2 reveals that the points allocated to net income are directionally higher under the *Disaggregated/Favorable* condition than the *Disaggregated/Unfavorable* condition, as revealed in Panel B of Table 2, there is not a significant *Favorable/Unfavorable* by *Forecast Aggregation* interaction ($p = 0.30$).

TABLE 2
Importance of Net Income to Participant Judgments

Panel A: Mean [standard deviation of] Importance of Net Income ^a		Favorable		Total
	Unfavorable			
Aggregated	26.43 [13.19] n = 49	26.18 [14.59] n = 50		26.30 [13.84] n = 99
Disaggregated	21.10 [11.43] n = 52	24.65 [12.92] n = 49		22.82 [12.25] n = 101
Column Mean	23.68 [12.54] n = 101	25.42 [13.74] n = 99		24.55 [13.15] n = 200

Panel B: Overall ANOVA		df	Mean Square	F-statistic	p-value ^b
Unfavorable/Favorable		1	136.73	0.80	0.37
Forecast Aggregation		1	587.79	3.44	0.03
Unfavorable/Favorable × Forecast Aggregation		1	180.91	1.06	0.30

Panel C: Mean [standard deviation of] Ratio of Points Assigned to Net Income by the Points Assigned to the Most Important Non-Net Income Item		Favorable		Total
	Unfavorable			
Aggregated	1.19 [0.76] n = 49	1.30 [1.35] n = 50		1.25 [1.09] n = 99
Disaggregated	0.91 [0.82] n = 52	1.16 [0.91] n = 49		1.03 [0.87] n = 101
Column Mean	1.05 [0.80] n = 101	1.23 [1.15] n = 99		24.55 [13.15] n = 200

(continued on next page)

Panel D: Overall ANOVA				
Source	df	Mean Square	F-statistic	p-value ^b
Unfavorable/Favorable	1	1.61	1.65	0.20
Forecast Aggregation	1	2.28	2.34	0.06
Unfavorable/Favorable × Forecast Aggregation	1	0.25	0.26	0.61

^a Participants allocated 100 points among the components of a hypothetical company's summary income statement (i.e., revenue, cost of sales, gross margin, research and development expense, selling, general, and administrative expenses, and net income) based on how important they believed each metric was for evaluating the company as a potential investment. The dependent measure is mean points allocated to net income. Greater earnings fixation is supported by greater points being allocated to net income. Participants received an *Aggregated (Disaggregated)* management forecast of earnings only (earnings and earnings components) and an earnings announcement. In addition, participants observed announced earnings that matched forecasted earnings, but whose revenues and R&D expense either *Unfavorable* or *Favorable* performance.

^b All p-values shown are two-tailed except for the *Forecast Aggregation* main effect whose p-value is one-tailed, given our directional prediction.

To supplement this evidence, we examine the spread of points allocated across net income and its components. If participants in the *Disaggregated* (*Aggregated*) forecast condition tend to consider net income more (less) as one of several inputs to evaluating the company, we expect the points they assign to the summary net income measure and to the most important net income component to be more (less) similar. Therefore, in an additional analysis, we calculate a ratio for each participant that divides the points assigned to the summary net income measure by the points assigned to the most important net income component (the component with the greatest number of assigned points). Our theory is supported if the calculated ratio is significantly greater in the *Aggregated* forecast condition than in the *Disaggregated* condition. The means of the calculated ratio are 1.25 and 1.03 (Table 2, Panel C) for the *Aggregated* and *Disaggregated* conditions, respectively, and are marginally different ($F = 2.34$, $p = 0.06$, one-tailed, Table 2, Panel D). These findings suggest that participants in the *Aggregated* condition assign 25 percent more points to the summary net income measure than the points they assign to the most important net income component, whereas participants in the *Disaggregated* condition assign a nearly equal number of points to net income and the most important net income component. Thus, participants in the *Disaggregated* condition appear to be more likely than participants in the *Aggregated* condition to consider net income as one of several similarly important inputs, further supporting our theory.¹⁷

Benchmark Framework as an Alternative Theoretical Explanation

Prior research in accounting has considered a benchmark framework when examining the impact of management forecasts on investor judgments (e.g., Krische 2005; Libby et al. 2006; Han and Tan 2007). An assumption underlying the benchmark framework as applied in investment settings is that investors' judgments might be sensitive to the number of benchmarks missed, met, or exceeded such that when a greater number of benchmarks are met (missed), investors perceive the company as a more (less) favorable potential investment (Han and Tan 2007). Because forecasts represent benchmarks of company performance, a benchmark framework represents an attractive approach to deriving predictions in our setting as an alternative to our theoretical framework.

In applying a benchmark framework to our experimental setting, one could plausibly reason that disaggregated forecasts highlight different benchmarks than aggregated forecasts. For example, participants in the *Disaggregated* forecast condition could use the forecasted line item amounts as benchmarks, while those in the *Aggregated* forecast condition could use prior year line item amounts as benchmarks because they observe no explicit forecasts to serve as benchmarks. More specifically, comparing the net number of benchmarks exceeded when prior year line item amounts are used as benchmarks to the net number of benchmarks exceeded when the forecasts of line items are used as benchmarks can yield a pattern of investment judgments similar to our findings.

Our experimental design does not allow us to discern clearly whether participants in our setting adopted a processing approach more consistent with our theoretical framework or a benchmark framework. Improving our understanding of the processing approach investors adopt when presented with forecasted information (and subsequent realizations) is important given the increasing frequency and importance of forecasts. Thus, parsing out these two theoretical perspectives represents a promising avenue for future research.

¹⁷ In addition, the proportion of participants who assigned their greatest number of points to Net Income was greater in the *Aggregated* condition (42 percent) than in the *Disaggregated* condition (33 percent). A one-sided Fisher Exact test (not tabulated) indicates that the difference in proportion is marginally significant ($p = 0.09$).

V. SUPPLEMENTAL ANALYSES

In this section, we examine whether the earnings-fixation-reducing effect of a disaggregated forecast is bounded by characteristics of the earnings announcement. We also provide evidence that our results are not fully explained by three plausible alternative explanations, namely, (1) perceived differences in information availability, (2) a pure management signal of the importance of certain financial statement items, or (3) perceived differences in forecast credibility across the level of forecast aggregation.

The Boundary Effects of Emphasis in Earnings Announcements

Although our study focuses on the ability of forecast disaggregation to reduce earnings fixation in investors' judgments, we include an examination of earnings announcement characteristics to examine boundary conditions for our hypothesis. Specifically, we consider whether an earnings announcement that emphasizes a bottom-line net income number can hinder the ability of a disaggregated forecast to reduce investors' earnings fixation. Investment professionals suggest that the extent to which a report emphasizes net income can affect investors' tendency to fixate on net income. For example, financial analysts assert that the power of a summary earnings number in proposing to eliminate any summary earnings from the income statement, in part, reduces "the focus on a single, arbitrary performance indicator, accounting net income" (CFA Centre for Financial Market Integrity 2007, 30).

We include an *Earnings Emphasis* manipulation in our experiment in which the earnings announcement either emphasizes net income by highlighting earnings and its percentage change in the press release heading, or deemphasizing net income by including in the press release heading the components of net income together with net income itself, and the respective percentage changes from the prior period.¹⁸ We then expand the ANCOVA model (untabulated) used to test our hypothesis to include all main and interactive effects of *Earnings Emphasis* on investors' judgments. Results indicate that whether the earnings announcement emphasizes or deemphasizes earnings has no impact on the ability of disaggregated forecasts to reduce earnings fixation. Specifically, an expanded ANCOVA (untabulated) reveals no three-way interaction between *Forecast Aggregation*, *Earnings Emphasis*, and *Unfavorable/Favorable* ($F = 0.26$, $p = 0.61$, two-tailed). Indeed, in observing the adjusted least square mean investor judgments, the disaggregated forecast appears to reduce earnings fixation when the earnings announcement emphasizes earnings (*Unfavorable* = 6.24 versus *Favorable* = 7.25) by nearly an identical amount as when the announcement deemphasizes earnings (*Unfavorable* = 5.76 versus *Favorable* = 6.76). Moreover, we find no significant interaction between the *Unfavorable/Favorable* and the *Earnings Emphasis* conditions ($F = 0.22$, $p = 0.64$, two-tailed) or the *Forecast Aggregation* and *Earnings Emphasis* conditions ($F = 1.07$, $p = 0.30$, two-tailed). Thus, earnings emphasis in the announcement does not appear to affect the reduction in fixation provided by disaggregating an earnings forecast.¹⁹

Differences in Available Information

We now consider the possibility that the effect of forecast disaggregation on reducing earnings fixation merely reflects a difference in the forecast information available to participants. As is

¹⁸ Using a Likert-type scale anchored on 1 = Not at all detailed and 11 = Very detailed, participants judged the detail of the emphasized announcement (5.44) and the deemphasized announcement (6.10) to be significantly different ($t = 2.21$, $p = 0.03$).

¹⁹ We had 24 participants take part in an experiment that was equal to our main experiment with the significant exception that no forecasts were given. Thus, this was a direct test of earnings emphasis alone. All 24 participants were placed in the *Unfavorable* condition and were randomly assigned to earnings announcements that either emphasized or deemphasized earnings. We found no significant effect for *Earnings Emphasis* in participant judgments ($F = 1.13$, $p = 0.30$, with initial judgments as the covariate) or points allocated to net income ($t = 0.86$, $p = 0.40$). Thus, in our setting, it does not appear that *Earnings Emphasis* alone eliminates (or even significantly reduces) earnings fixation.

inherently the case in natural settings (and following Hirst et al. [2007]), disaggregated forecasts in our experiment specify a greater number of items than do aggregated forecasts. Thus, one could argue that participants who observed aggregated forecasts ignored the net income components because the lack of forecast information made it impossible to assess the components as positive or negative news. This is likely not the case, however. First, if participants in the *Aggregated* condition were unable to evaluate the net income components, then we would expect some of them to assign approximately zero points to the net income components. The results, in contrast, show that all participants assigned some points to the net income components. Second, we chose the values used in the disaggregated forecast to represent an extension of the linear trend of the 2005 and 2006 income statement amounts. This should allow participants in the *Aggregated* condition to derive an expectation for the components of net income similar to the expectations derived by those in the *Disaggregated* condition.²⁰ Additional data below suggest that participants in both aggregation levels derive similar expectations.

An investor's surprise following a company announcement represents the extent to which reported amounts deviate from the investor's expectations. Therefore, if participants in the *Aggregated* condition did not form expectations of revenue, research and development, and net income (because they received no forecasts for earnings components), then we would expect them to be less surprised at the amounts reported for these items in the earnings announcement than participants in the *Disaggregated* condition (who received forecasts of the components). To examine this proposition, we ask participants the following question with regard to revenue, research and development, and net income: "How surprised were you by the amount of Revenues/R&D/Net Income reported for 2007?" Participants answered on a Likert-type scale anchored on 1 = "Not at All Surprised" and 11 = "Very Surprised." The mean surprise for announced net income is 5.52 and 5.11 for those in the *Aggregated* and *Disaggregated* forecast conditions, respectively. The mean surprise for announced revenue is 5.47 and 5.82 for those in the *Aggregated* and *Disaggregated* forecast conditions, respectively. And the mean surprise for announced R&D is 6.24 and 6.52 for those in the *Aggregated* and *Disaggregated* forecast conditions, respectively. An ANOVA (not tabulated) reveals no significant main effect for *Forecast Aggregation* (all p-values > 0.30), nor any significant interaction effects (all p-values > 0.49) for any of these dependent measures.²¹ Thus, our findings do not seem to be the result of participants who received an aggregated forecast not developing expectations for the components of net income.²²

Pure Management Signal of Importance of Certain Financial Statement Items

We next consider whether our observed reduction in fixation is driven purely by participants' perception that the level of forecast aggregation is a signal from management regarding the importance of certain financial statement items. In natural settings, companies can be selective about what items to include in the forecast (see Exhibit 1). In such cases, it is reasonable for

²⁰ Our design choice of a linear trend was made in an attempt to maximize internal validity. Of course, in the natural setting, forecasts of income statement line items do not always match the prior linear trends of those line items. However, we see no reason why our inferences would not extend to such a setting.

²¹ Importantly, the level of surprise expressed by participants in the *Unfavorable* condition significantly differed from that expressed by participants in the *Favorable* condition for revenues ($p = 0.02$), R&D ($p < 0.01$), and net income ($p = 0.07$). This indicates that our null result for aggregation was not due to our measure for surprise being inadequately sensitive to detect differences in participants' levels of surprise.

²² To confirm that participants across aggregation levels attended to the amounts reported in the earnings announcement, we asked participants to recall the amounts reported for 2008 revenue, R&D Expense, and net income. We also asked them to recall how much 2007 amounts for these line items changed in 2008. There were no differences across *Forecast Aggregation* conditions for any of these recall measures (all p-values > 0.10).

investors to conclude that managers might use a disaggregated forecast to highlight those items that are particularly relevant and exclude items that are less relevant. However, our experimental design makes such signaling less likely; recall that our disaggregated forecast includes *all* items reported in the subsequent earnings announcement. The apparent lack of management selectivity should reduce the likelihood that participants perceived the disaggregation as a signal of the particular items that were most important.

Evidence from our data also runs counter to a signaling explanation for our findings. Recall that, in addition to manipulating the aggregation level of forecasts, we also manipulated whether a subsequent earnings announcement emphasized net income. Emphasizing net income or net income components in an announcement would suggest a similar management signal to any signal ascribed to the aggregation level of a forecast. However, as noted previously, we find no evidence that emphasizing earnings in an earnings announcement reduces earnings fixation. Thus, a strict signaling explanation for our overall findings is unlikely.

Forecast Credibility

Hirst et al. (2007) find that forecast aggregation affects investors' judgments of forecast credibility. Therefore, we also consider whether our results can be explained by differences in forecast aggregation causing differences in perceived forecast credibility.²³ Similar to Hirst et al. (2007) we ask participants to judge the credibility of the forecast provided. Participants provided responses on a Likert-type scale with end points, 1 = Not at All Credible and 11 = Very Credible. It should be noted that Hirst et al. (2007) elicit participants' judgments of forecast credibility directly after observing a management forecast (and the participants never observe a subsequent earnings announcement). In contrast, we do not elicit judgments of forecast credibility until after participants observe both the management forecast and the subsequent earnings announcement (so as not to explicitly cue participants to think about forecast credibility before viewing the subsequent earnings announcement and making their final investment judgments).

Our participants' mean ratings of forecast credibility are higher when they observe a disaggregated forecast (mean = 7.15) as compared to an aggregated forecast (mean = 7.00), directionally consistent with the results of Hirst et al. (2007). However, an ANOVA (not tabulated) reveals no main or interactive effects when forecast credibility is the dependent measure and our three independent variables and all of their interactions are included. Thus, it does not appear that differences in participants' judgments of forecast credibility are driving our results. In addition, it is not surprising that our results for the effect of forecast aggregation on participants' judgments of forecast credibility are not as strong as those reported in Hirst et al. (2007), as our participants viewed a realization of the forecast before evaluating forecast credibility.

²³ In addition to the credibility question described below, we asked related questions about management's trustworthiness, competence, and credibility; questions about management's opportunities, motives, and incentives to manage earnings; questions about how constrained management was to manage earnings; and a question about the quality of reported net income. No significant differences across aggregation levels were noted for any of the measures (all p-values > 0.15, two-tailed). However, participants' responses did differ across the *Unfavorable* and *Favorable* conditions, as expected. Specifically, participants in the *Unfavorable* condition who observed a cut in R&D (as opposed to an increase in R&D) rated management as less trustworthy ($p = 0.08$, one-tailed), less forthcoming ($p = 0.01$, one-tailed) and less competent ($p = 0.01$, one-tailed) than those in the *Favorable* condition. In addition, participants in the *Unfavorable* condition rated the quality of net income lower than those in the *Favorable* condition ($p = 0.06$, one-tailed). These latter results suggest that participants did attend to management characteristics and the related quality of net income associated with cutting R&D (as opposed to increasing R&D) to arrive at a forecasted realization; however, our participants did not believe that these characteristics differed by the level of forecast aggregation. Finally, it is unclear, *ex ante*, if participants' responses to the questions about management's opportunities, motives, and incentives to manage earnings, and how constrained management was to manage earnings would differ across the *Unfavorable* and *Favorable* conditions. In fact, no significant differences were observed (all p-values > 0.10, two-tailed).

VI. CONCLUSION

This study examines a mechanism intended to reduce investors' susceptibility to earnings fixation (i.e., excessive reliance on accounting earnings as the primary input in investment judgments, without fully considering other information that may be relevant to evaluating the company's investment prospects). We find that investors are less susceptible to earnings fixation when they initially observe a disaggregated management forecast (which forecasts earnings and its components) than when they observe an aggregated forecast (which forecasts only earnings).

Specifically, participants' investment judgments reflect less earnings fixation when they initially observe a disaggregated management earnings forecast. This reduction in earnings fixation appears to be associated with investors interpreting the summary net income figure as one of several similarly important evaluation inputs rather than a substantially more important input (relative to its components). We provide additional evidence that our results cannot be fully explained by several plausible alternative explanations: (1) perceived differences in information availability across levels of aggregation, (2) perceived management signals of the importance of certain financial statement items, or (3) perceived differences in forecast credibility across levels of aggregation. Finally, we provide evidence that suggests our results are not bounded by the level of emphasis on net income in the subsequent earnings announcement.

As with all experimental research, our study is not without limitations. Participants in our study did not have access to all the information they might have available in a natural setting. Moreover, a limitation specific to our study is that an aggregated (disaggregated) forecast, by definition, provides investors with a single (multiple) benchmark(s). Thus, the level of forecast aggregation, in our experiment as well as in the real world, is inexorably linked with the number of benchmarks in the forecast, forming a natural environmental confound. In addition, while we report evidence that a reduction in earnings fixation results from a disaggregated forecast activating an investor-held knowledge structure, we cannot determine whether this activation occurs prior to or subsequent to investors observing the earnings announcement. Finally, because managers may strategically issue a disaggregated forecast to influence investor perceptions of firm value, providing disaggregated forecasts may not always lead investors to optimal judgments.

Despite these limitations, our results provide important insights regarding a potential mechanism to reduce investors' susceptibility to earnings fixation—an institutional phenomenon that can be undesirable to managers and investors. In addition, our study extends the disclosure literature in accounting by providing evidence that a forecast (and its characteristics) can influence how investors interpret subsequently reported accounting information. Our findings should be of interest to managers, who have incentives to reduce pressures for myopic behavior, investors, whose investment returns may be adversely affected by such myopic behavior, and regulators, who have expressed concern about investors' susceptibility to earnings fixation and the resulting impact on manager behavior.

APPENDIX A

DISAGGREGATED MANAGEMENT FORECAST²⁴

LearningWare Provides Forecast for 2007

Tallahassee, FL—LearningWare today provided a forecast for the full year, which ends on December 31, 2007. LearningWare expects revenues to increase 20 percent to \$43.2 million. The gross profit margin is projected to remain at 49 percent. Research and development expenses are

²⁴ The forecast shown is for the *Disaggregated* condition. Participants in the *Aggregated* condition received only the last sentence of the text and received only the audited 2005 and 2006 summary income statements.

expected to increase 19 percent to \$6.9 million. Selling, general, and administrative expenses are expected to increase 20 percent to \$4.9 million. LearningWare expects 2007 net income to increase 22 percent to \$6.1 million.

Income Statement (all figures in millions)

Fiscal Year ending December 31

	Audited		Unaudited
	2005	2006	2007 (Forecast)
Revenue	\$30.0	\$36.0	\$43.2
Cost of Sales	15.3	18.4	22.0
Gross Profit	14.7	17.6	21.2
Research and Development Expense	4.9	5.8	6.9
Selling, General, and Administrative Expense	3.4	4.1	4.9
Earnings Before Income Taxes	6.4	7.7	9.4
Income Tax Expense	2.2	2.7	3.3
Net Income	\$4.2	\$5.0	\$6.1

About LearningWare: LearningWare (<http://www.learningware.com>) develops and sells various software applications that help schools educate and evaluate students and help businesses assess the skills of their employees. This news release contains forecasted information that is not audited.

APPENDIX B

UNFAVORABLE/FAVORABLE ANNOUNCED INCOME STATEMENTS²⁵

Panel A: Unfavorable Condition

	Audited		
	2005	2006	2007
Revenue	\$30.0	\$36.0	\$39.6
Cost of Sales	15.3	18.4	20.2
Gross Profit	14.7	17.6	19.4
Research and Development Expense	4.9	5.8	5.1
Selling, General, and Administrative Expense	3.4	4.1	4.9
Earnings Before Income Taxes	6.4	7.7	9.4
Income Tax Expense	2.2	2.7	3.3
Net Income	\$4.2	\$5.0	\$6.1

(continued on next page)

²⁵ Panels A and B indicate the summary income statements participants observed when in the *Unfavorable* and *Favorable* announcement conditions, respectively. Amounts highlighted (were not highlighted in actual experiment) indicate the fundamental differences between the two conditions although net income is identical. Earnings fixation is reduced (enhanced) to the extent that participants' investment judgments discern (do not discern) between the *Unfavorable* and *Favorable* income statements.

Panel B: Favorable Condition

	Audited		
	2005	2006	2007
Revenue	\$30.0	\$36.0	\$46.8
Cost of Sales	15.3	18.4	23.9
Gross Profit	14.7	17.6	22.9
Research and Development Expense	4.9	5.8	8.6
Selling, General, and Administrative Expense	3.4	4.1	4.9
Earnings Before Income Taxes	6.4	7.7	9.4
Income Tax Expense	2.2	2.7	3.3
Net Income	\$4.2	\$5.0	\$6.1

REFERENCES

- Ajinkya, B. B., and M. J. Gift. 1984. Corporate managers' earnings forecasts and symmetrical adjustments of market expectations. *Journal of Accounting Research* 22 (2): 425–444.
- Barton, J., and M. Mercer. 2005. To blame or not to blame: Analysts' reactions to external explanations for poor financial performance. *Journal of Accounting and Economics* 39 (3): 509–533.
- Bowen, R. M., X. Chen, and Q. Cheng. 2008. Analyst coverage and the cost of raising equity capital: Evidence from underpricing of seasoned equity offerings. *Contemporary Accounting Research* 25 (3): 657–699.
- Bushee, B. J. 1998. The influence of institutional investors on myopic R&D investment behavior. *The Accounting Review* 73 (3): 305–333.
- . 2001. Do institutional investors prefer near-term earnings over long-run value? *Contemporary Accounting Research* 18 (2): 207–246.
- CFA Centre for Financial Market Integrity. 2006. Breaking the short-term cycle. *CFA Institute Centre Publications*: 1–20. Available at: <http://www.cfapubs.org/doi/pdf/10.2469/ccb.v2006.n1.4194>.
- . 2007. A comprehensive business reporting model: Financial reporting for investors. *CFA Institute Centre Publications*: 1–59. Available at: <http://www.cfapubs.org/doi/abs/10.2469/ccb.v2007.n6.4818>.
- . 2008. Short-termism survey: Practices and preferences of investment professionals. *CFA Institute Centre Publications*: 1–33. Available at: http://www.cfainstitute.org/Survey/short_termism_survey_report_may_2008.pdf.
- Chen, S., D. Matsumoto, and S. Rajgopal. Forthcoming. Is silence golden? An empirical analysis of firms that stop giving quarterly earnings guidance. *Journal of Accounting & Economics*.
- Cheng, M., K. R. Subramanyam, and Y. Zhang. 2007. Earnings guidance and managerial myopia. Working paper, University of Southern California.
- Cotter, J., A. I. Tuna, and P. D. Wysocki. 2006. Expectations management and beatable targets: How do analysts react to explicit earnings guidance? *Contemporary Accounting Research* 23 (3): 593–624.
- Economist, The*. 2006. The sound of silence. *The Economist* (April): 29.
- Elliott, W. B., F. D. Hodge, J. J. Kennedy, and M. Pronk. 2007. Are M.B.A. students a good proxy for nonprofessional investors? *The Accounting Review* 82 (1): 139–168.
- , ———, and K. E. Jackson. 2008. The association between nonprofessional investors' information choices and their portfolio returns: The importance of investing experience. *Contemporary Accounting Research* 25 (2): 473–498.
- Fischhoff, B. 1977. Perceived informativeness of facts. *Journal of Experimental Psychology. Human Perception and Performance* 3 (2): 349–358.
- Graham, J. R., C. R. Harvey, and S. Rajgopal. 2005. The economic implications of corporate financial reporting. *Journal of Accounting and Economics* 40 (1–3): 3–73.

- Han, J., and H.-T. Tan. 2007. Investors' reactions to management guidance forms: The influence of multiple benchmarks. *The Accounting Review* 82 (2): 521–543.
- Herz, R. H. 2006. Testimony of Robert H. Herz, Chairman, Financial Accounting Standards Board before the Capital Markets, Insurance and Government Sponsored Enterprises Subcommittee of the Committee on Financial Services. March 29. Available at: http://www.fasb.org/cs/ContentServer?c=Document_C&pagename=FASB%2FDocument_C%2FDocumentPage&cid=1176156900579.
- Higgins, E. T., and G. A. King. 1981. Accessibility of social constructs: Information processing consequences of individual and contextual variability. In *Personality, Cognition, and Social Interaction*, edited by Kihlstrom, J. F., and N. Cantor. Hillsdale, NJ: L. Erlbaum Associates.
- Hirst, D. E., L. Koonce, and S. Venkataraman. 2007. How disaggregation enhances the credibility of management earnings forecasts. *Journal of Accounting Research* 45 (4): 811–837.
- Houston, J. F., B. I. Lev, and J. Tucker. 2010. To guide or not to guide? Causes and consequences of stopping quarterly earnings guidance. *Contemporary Accounting Research* 27 (1).
- Hutton, A. P., and P. Stocken. 2009. Prior forecasting accuracy and investor reaction to management earnings forecasts. Working paper, Boston College and Dartmouth College.
- Investment Company Institute and the Securities Industry and Financial Markets Association (ICI/SIFMA). 2008. Equity and bond ownership in America. Available at: http://www.ici.org/pdf/rpt_08_equity_owners.pdf.
- Janjigian, V., and M. Ozanian. 2006. Gimmie guidance. *Forbes.com*.
- Johnson, M. F., R. Kasznik, and K. K. Nelson. 2001. The impact of securities litigation reform on the disclosure of forward-looking information by high technology firms. *Journal of Accounting Research* 39 (2): 297–327.
- Krische, S. D. 2005. Investors' evaluations of strategic prior-period benchmark disclosures in earnings announcements. *The Accounting Review* 80 (1): 243–268.
- Lansford, B., B. Lev, and J. W. Tucker. 2010. Causes and consequences of disaggregating earnings guidance. Working paper, Northwestern University, New York University, and University of Florida.
- Libby, R., R. Bloomfield, and M. Nelson. 2002. Experimental research in financial accounting. *Accounting, Organizations and Society* 27: 777–812.
- , and H.-T. Tan. 1999. Analysts' reactions to warnings of negative earnings surprises. *Journal of Accounting Research* 37 (2): 415–435.
- , ———, and J. E. Hunton. 2006. Does the form of management's earnings guidance affect analysts' earnings forecasts? *The Accounting Review* 81 (1): 207–225.
- McCafferty, J. 2007. Companies are trying to shift investors' focus from short-term results to long-term goals. *CFO Magazine* (May 1).
- Murray, D. 1991. Data fixation: Methodological refinements and additional empirical evidence. *Behavioral Research in Accounting* 3: 25–38.
- Penman, S. H. 2007. *Financial Statement Analysis and Security Valuation*. New York, NY: McGraw-Hill/Irwin.
- Schrand, C. M., and B. R. Walther. 2000. Strategic benchmarks in earnings announcements: The selective disclosure of prior-period earnings components. *The Accounting Review* 75 (2): 151–177.
- Skinner, D. J. 1994. Why firms voluntarily disclose bad news. *Journal of Accounting Research* 32 (1): 38–60.
- . 1997. Earnings disclosures and stockholder lawsuits. *Journal of Accounting and Economics* 23 (3): 249–282.
- Tan, H.-T., R. Libby, and J. E. Hunton. 2002. Analysts' reaction to earnings preannouncement strategies. *Journal of Accounting Research* 40 (1): 223–246.
- Taylor, S. E., and S. T. Fiske. 1979. Salience, attention, and attribution: Top of the head phenomena. In *Advances in Experimental Social Psychology*, edited by Berkowitz, L. New York, NY: Academic Press.
- Tversky, A., and D. Kahneman. 1973. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5 (2): 207–232.
- , and ———. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185 (4157): 1124–1131.
- Wyer, R. S., and T. K. Srull. 1981. Category accessibility: Some theoretical and empirical issues concerning

- the processing of social stimulus information. In *Social Cognition: The Ontario Symposium*, edited by E. T. Higgins, C. P. Herman, and M. P. Zanna, 161–197. Hillsdale, NJ: Erlbaum.
- Yi, Y. 1990a. Cognitive and affective priming of the context for print advertisements. *Journal of Advertising* 19 (2): 40–48.
- . 1990b. The effects of contextual priming in print advertisements. *The Journal of Consumer Research* 17 (2): 215–222.

Percent Accruals

Nader Hafzalla
Russell Lundholm
University of Michigan
E. Matthew Van Winkle
Voyant Advisors, LLC

ABSTRACT: We document how the effectiveness of an accruals-based trading strategy changes with the benchmark used to identify an extreme accrual. We measure “percent accruals” as accruals scaled by earnings, rather than total assets, and show that this seemingly small change produces a radically different sort of the data. We find that a trading strategy based on percent accruals yields significantly larger annual hedge returns than the traditional accruals measure, and does so mostly by improving the long position in low-accrual stocks. The hedge returns are also significant in all but the lowest quintile of arbitrage risk. We show that percent accruals more effectively select firms where the difference between sophisticated and naïve forecasts are the most extreme. As such, our results are consistent with the earnings fixation hypothesis and are inconsistent with some alternative explanations for the accrual anomaly.

Keywords: *accruals; market inefficiency; abnormal returns.*

Data Availability: *Data used in this study are available from the sources identified in the text.*

I. INTRODUCTION

The accruals anomaly, originally identified by Sloan (1996), is one of the most important findings in accounting research. By purchasing firms with low accruals and selling short firms with high accruals, Sloan (1996) documents a significant hedge return. A large literature has followed exploring the cause of this market inefficiency, but also questioning the size and persistence of the anomaly. In the vast majority of these studies, accruals are measured as earnings less cash from operations, scaled by average total assets. In this study we propose a small change in the definition of accruals that yields surprisingly different results. Instead of scaling

We thank Rachel Hayes, Matthew Magilke, Richard Sloan, Teri Yohn, and the workshop participants at Barclay’s Global Investor, University of British Columbia, University of Massachusetts, University of Michigan, The Pennsylvania State University Research Conference, The University of Utah, and Washington University for helpful comments.

Nader Hafzalla died June 29, 2006 while in the fourth year of his Ph.D. program at the University of Michigan. He was a talented scholar and a terrific friend. This work is based on two separate projects that Nader was involved in and is dedicated to his memory.

Editor’s note: Accepted by Steven Kachelmeier, with thanks to Dan Dhaliwal for serving as editor on a previous version.

Submitted: February 2008

Accepted: May 2010

Published Online: January 2011

accruals by average total assets, we scale by earnings. We label this definition of accruals as “percent accruals” and argue that it is a more natural interpretation of the original idea that investors are fixated on reported earnings and do not distinguish between accruals and cash flows.

Percent accruals identify an extreme observation as one for which the accrual component makes up a large positive or negative fraction of the total earnings for that year.¹ This measure of accruals focuses on the composition of earnings, distinguishing how much is cash and how much is accrual. We show that this seemingly simple change in the definition of an extreme accrual produces a radically different sort of the data; only 12 percent of the data in the first decile of the traditional accruals measure are in the first decile of our new measure. By simply changing the benchmark for an extreme accrual, we find a hedge return that is much larger than the return based on the traditional definition of accruals, and is particularly successful for the long position.

The percent accruals measure differs most from the traditional accruals measure in the lowest decile—the firms for which a trading strategy takes a long position. The returns to this part of a hedge trading strategy are particularly important because the transaction costs and other limits to arbitrage are significantly lower on a long position than on a short position. We find a 5.53 percent abnormal return to the long position in the first decile of percent operating accruals that is significant at the 0.001 level. This return is four times as large as the 1.27 percent abnormal return in the first decile of the traditional accruals measure, which is not significantly different from zero.² Further, the firms in the first decile of percent operating accruals have an average market value of \$1,510 million, as compared to only \$474 million for the first decile of the traditional operating accruals measure. This suggests that the transaction costs to trading on a percent accruals anomaly are probably much lower than for the traditional accruals measure, and make the evidence for inefficient pricing of percent accruals even more compelling. In addition, we show that the hedge returns based on percent accruals are significant in all but the lowest quintile of arbitrage risk, making the results even harder to reconcile with an efficient market.

Because *percent accruals*—accruals measured as a percent of absolute income—is a new definition of accruals, we examine what types of firms end up in the extreme high and low portfolios of this measure, and why they might experience excess positive or negative returns in the future. The typical firm in the extreme negative percent operating accrual portfolio has large positive cash from operations, but then accrues income back down to something much closer to zero; similarly, to end up in the extreme positive percent operating accrual portfolio, a typical firm has large negative cash from operations, but accrues income up to something near zero. We show that these extreme combinations of cash from operations and accruals are exactly the combinations that produce the most extreme differences between a sophisticated income forecast—one that distinguishes between cash flows and accruals—and a naïve forecast. As a result, the percent accruals measure is arguably superior to the traditional accruals measure at identifying cases for which the earnings fixation hypothesis predicts the largest mispricing.

The percent accruals measure has other advantages. We show that the extreme percent accruals are not disproportionately sensitive to special items, as Dechow and Ge (2006) report for traditional operating accruals, and the hedge returns are equally large in the subsample with or without special items. We also show that the long position is more effective in the subsample of

¹ In order to assure that the sign of the accrual is maintained when we sort firms into portfolios, we scale by the absolute value of net income. Because net income can be at or near zero, the result will sometimes be extreme. Note, however, that we only use this variable to sort firms into ten portfolios; beyond placing a firm into the first or tenth portfolio, the extremity of the variable is irrelevant.

² As discussed in the next section, comparing the hedge returns from different accruals studies is difficult due to a prevalent look-ahead bias documented in Kraft et al. (2006).

firms with losses than those with gains, unlike the traditional accrual measure that has been shown to have no predictive power for future returns when the sample is limited to loss firms (Dopuch et al. 2009).

Aside from offering new evidence consistent with the accruals anomaly, our study contributes to the growing literature that attempts to explain the cause of the anomaly. The main question in this literature is whether the anomaly is driven by investors' failure to understand the different mean-reversion tendencies of cash flows versus accruals—commonly known as the earnings fixation hypothesis. One inconsistency between prior accrual anomaly results and this hypothesis is that existing studies find very little evidence of mispricing in the low-accrual portfolio.³ An alternative to the earnings fixation hypothesis is offered by Kothari et al. (2006). They argue that managers of overpriced stocks manipulate earnings up to maintain the overvaluation for as long as possible, the market is fooled by these positive accruals, and this causes the high-accrual portfolio to earn a significant negative return when the accruals reverse and the market corrects itself. In contrast, managers of undervalued firms have no incentive to manipulate earnings down, and consequently there is no predicted undervaluation for the low-accrual portfolio, consistent with the near-zero returns observed in this portfolio. Our results reinstate the significant underpricing of the low-accrual portfolio when accruals are measured relative to earnings instead of total assets. This is inconsistent with the asymmetric prediction of Kothari et al.'s (2006) agency hypothesis. More generally, since neither the earnings fixation hypothesis nor the agency hypothesis is conditioned on exactly how accruals are scaled, our results suggest that a more refined theory is in order.

Another competitor to the earnings fixation hypothesis is offered by Fairfield et al. (2003), who note that when accruals are scaled by assets, the resulting measure is essentially the percentage change in operating assets. If investors fail to understand that investments tend to have diminishing returns, then they may overprice large increases in assets and underprice large decreases in assets. Rather than fixating on earnings, investors simply fail to understand the economics of investment. Our results do not invalidate this interpretation of the past accruals literature or the evidence supporting the "growth" hypothesis, but this interpretation does not apply to our rescaled measure of accruals. Finally, as discussed above, the percent accruals measure produces a better implementation of the trading strategy predicted by the earnings fixation hypothesis than the traditional accruals measure, and it generates superior excess returns, lending further evidence in support of the earnings fixation hypothesis.

Two other studies are related to ours. Dopuch et al. (2009) show that the traditional accrual anomaly holds only for profit firms; the hedge return in the loss firm sample is not statistically different from zero. This finding is due primarily to loss firms with large negative accruals; this portfolio is supposed to have positive excess returns in the accrual anomaly, yet the returns are -0.2 percent in the loss sample. Firm-years with losses comprise roughly one third of the sample, so this is not an inconsequential limitation. In contrast, the return to a hedge strategy based on percent accruals is slightly larger in the loss subsample than in the gain subsample, and is significant in both samples.

Cheng and Thomas (2005) assess 22 different abnormal accruals models and 5 different design choices as a comprehensive investigation into the specifications used to study the accruals anomaly. They find significant variation in the size of the anomaly over the 704 permutations considered. However, all of their accruals models are scaled by some measure of firm size, either total assets or market value, but never by earnings.

³ See Kraft et al. (2006) and Zach (2004) specifically for this topic, as well as Beaver et al. (2007), Beneish and Vargus (2002), D'Avolio et al. (2002), Houge and Loughran (2000), and Lesmond and Wang (2006).

In the next section we describe our new definition of extreme accruals, in Section III we discuss the sample properties of the different accrual measures, and in Section IV we present stock return results based on our new measures. We investigate the cause of the percent accruals' success in Section V and conclude in Section VI.

II. IDENTIFYING EXTREME ACCRUALS

To begin we define the two "traditional" accrual measures found in the literature. The most common definition is:

$$\text{Traditional Operating Accruals} = (\text{Net Income} - \text{Cash from Operations}) / \text{Average Total Assets}.$$

Sloan's (1996) original definition was based on balance sheet changes but, starting with Collins and Hribar (2002), the data from the statement of cash flows have become the preferred definition because they exclude the effects of acquisitions and foreign currency translation adjustments. We use the same variables as in Kraft et al. (2006), calculating traditional operating accruals as the difference between net income (Compustat item 172) and cash from operations (item 308), then dividing by the average of total assets (item 6).

Recently, Richardson et al. (2005) introduced a more comprehensive measure of accruals, specified as:

$$\begin{aligned} \text{Traditional Total Accruals} = & [\text{Net Income} \\ & - (\text{Net Dividends and Distributions to/from Equityholders} \\ & + \text{increase in the cash balance})] / \text{Average Total Assets}. \end{aligned}$$

Traditional total accruals are also computed using the cash flow statement, as in Richardson et al. (2005), calculated here as the difference between income before extraordinary items (item 172) less total operating, investing, and financing cash flows (items 308, 311, and 313) plus sales of common stock (item 108) less stock repurchases and dividends (items 115 and 127), divided by the average of total assets (item 6).

Note that both definitions of accruals, and all those we could find in the literature (see Cheng and Thomas 2005), scale by some measure of firm size, in this case total assets. So firms in the first decile of accruals have large negative accruals relative to their size. Another natural benchmark for extremity would be relative to the firm's earnings for the period. By computing accruals as a percent of earnings, the measure focuses on the *composition* of earnings—the relative amounts of cash and accruals. To assure that negative accruals are in the lower portfolios and positive accruals are in higher portfolios, we scale by the absolute value of earnings. That is, our percent accruals measures are:

$$\text{Percent Operating Accruals} = (\text{Net Income} - \text{Cash from Operations}) / |\text{Net Income}|$$

and

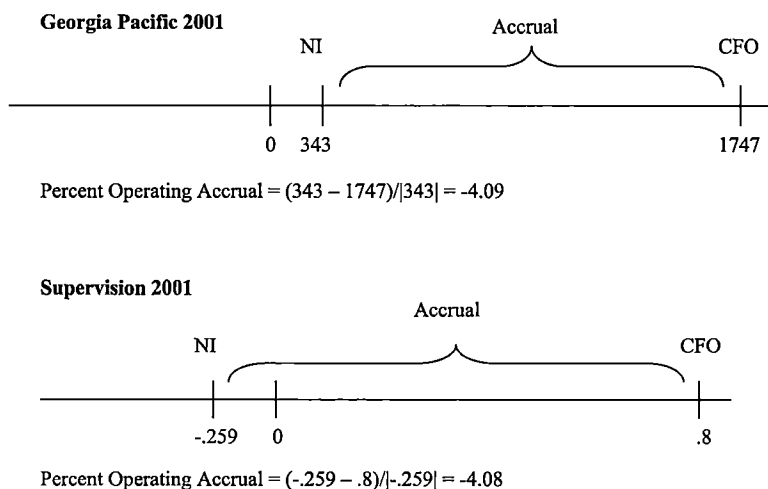
$$\begin{aligned} \text{Percent Total Accruals} = & [\text{Net Income} \\ & - (\text{Net Dividends and Distributions to/from Equityholders} \\ & + \text{increase in the cash balance})] / |\text{Net Income}|. \end{aligned}$$

For the percent accruals measures, firms in the first accruals decile have large negative accruals and small positive or small negative earnings; this combination yields a large negative numerator and a small denominator. The cutoff to be in the lowest decile of percent operating

accruals varies by year, ranging from -5.7 to -3.1 . This implies that all the firms in this decile have positive cash from operations and their negative accrual pushes their net income back toward zero (if cash from operations was negative, then a negative accrual would only push net income further from zero, making the denominator greater than the numerator and the resulting value greater than -1). Similarly, firms in the highest decile of percent accruals tend to have negative cash from operations and their positive accrual pushes their net income back toward zero. Because the denominator is the absolute value of net income, and net income can be near zero, some values of the percent accruals are extreme.⁴ Note, however, that this variable is only used to sort firms into portfolios; extreme values have no more impact on the results than do the more moderate values within the same portfolio.

Consider the examples of Georgia Pacific and Supervision, as shown in Figure 1, which are taken from the lowest decile of percent operating accruals. Both examples illustrate how a large positive cash flow and a large negative accrual combine to yield an extreme negative value of percent operating accruals. These examples also illustrate why the percent operating accruals measure is insensitive to firm size, unlike the traditional operating accruals measure: Georgia Pacific has earnings and cash flows (and unreported total assets) that are orders of magnitude greater than Supervision's numbers, yet they have almost identical percent operating accruals. As we show later, this results in significantly larger firms in the first decile of percent operating accruals than in the first decile of traditional operating accruals. These examples also illustrate why it makes sense to use the absolute value of earnings in the denominator, as one would not want to treat these two very similar examples differently. Finally, these examples illustrate why very poorly performing firms are unlikely to be sorted into the first decile of a percent accruals measure. Firms in the lowest percent operating accruals decile have positive cash flows and

FIGURE 1
Examples from the First Decile of Percent Operating Accruals



⁴ Five observations have zero net income. We sorted these into either decile 1 or decile 10, depending on the sign of the accrual.

earnings near zero, while firms in the lowest traditional operating accruals decile have large losses and large negative cash flows. Consequently, sorting firms by percent accruals measures will spread poor performing observations across the distribution, rather than concentrate them in the lowest decile. We document these empirical regularities in the next section.

III. SAMPLE AND DESCRIPTIVE STATISTICS

We use the merged 2008 Compustat/CRSP database on the WRDS system to collect all firms with available data to construct our accrual measures. Consistent with prior studies, we exclude financial firms. The only data requirements are that the firm has sufficient financial statement data on Compustat to compute traditional operating accruals in the current year, and it has return data available at the portfolio formation date (the first day of the fourth month after the fiscal year-end). The result is a sample of 81,526 firm-years spanning the years 1989–2008. As discussed later, we use the prior year's decile breaks to form portfolios, resulting in 19 years of annual portfolio returns for our returns tests. Table 1 reports descriptive statistics. Because our data requirements are so minimal, the sample is representative of the nonfinancial market as a whole for this time period.

TABLE 1
Descriptive Statistics

Variable	Mean	Median	Standard Deviation	Lower Quartile	Upper Quartile
Percent operating accruals	−1.6860	−0.6753	5.2630	−1.6040	−0.0644
Traditional operating accruals	−0.0670	−0.0517	0.1345	−0.1077	−0.0059
Percent total accruals	−0.8576	0.1737	5.2010	−1.0090	1.0000
Traditional total accruals	−0.0730	0.0084	0.2784	−0.0986	0.0668
Return on Assets	−0.0326	0.0311	0.2235	−0.0503	0.0764
Cash Return on Assets	0.0349	0.0685	0.1765	−0.0062	0.1285
Total Cash Flow Return on Assets	−0.1077	−0.0017	0.2875	−0.0944	0.0341
Market Value of Equity	1,696	138	5,449	30	743
Book Value/Market Value	0.6310	0.4949	0.6239	0.2698	0.8236
Price Per Share	16.61	10.72	17.16	3.88	24.00
Three-Year Sales Growth	0.1076	0.0589	0.4795	−0.1025	0.1869
Proportion of firms with losses	0.3445	1.0000	0.4752	0.0000	1.0000
% of Firms with Special Item <0	0.1988	0.0000	0.3991	0.0000	0.0000
Arbitrage Risk	0.1516	0.1331	0.0840	0.0912	0.1892

The sample period is 1988–2008, consisting of 81,526 firm-years. Traditional operating accruals are defined as Net Income (Compustat item 172) less Cash from Operations (item 308) divided by average total assets (item 6). Traditional total accruals are computed as Net Income less Net Dividends/Distributions (item 115 + item 127 − item 108) and the change in cash (item 308 + item 311 + item 313) divided by average total assets. Percent Operating Accruals and Percent Total Accruals have the same numerators as Operating Accruals and Total Accruals, respectively, but the denominator for each is the absolute value of Net Income. Return on Assets is Net Income divided by average total assets. Cash Return on Assets is cash from operations divided by average total assets. Total Cash Flow Return on Assets are Net Dividends/Distributions plus the change in cash divided by average total assets. Market Value of Equity is measured at the fiscal year-end (item 25 times item 199). Price (item 199) and Book Value (item 60) are measured at the fiscal year-end. Three-year Sales growth is the average annual sales growth in the current year and prior two years. Special items (item 17) are coded as 0 if missing. Arbitrage Risk is measured as the standard deviation of the residuals from a regression of firm-specific returns on the CRSP equally weighted index for up to 48 months preceding the month of return portfolio formation. Amounts are Winsorized at 1 percent and 99 percent for this table and Table 2 only.

Table 2 gives the distribution of the four accruals measures as well as the distributions of other firm characteristics across deciles. The first row in Panel A shows that to be in the lowest decile of percent operating accruals in the average year, the negative accrual needs to be at least 4.04 times earnings; to be in the highest decile, the positive accrual needs to be at least 0.82 times earnings. Comparing Panel A to the traditional operating accruals in Panel B shows that the types of firms that are sorted into the extreme deciles are quite different. The mean market value in decile 1 of percent operating accruals is \$1,510 million versus \$474 million for traditional operating accruals. The firms in decile 1 of percent operating accruals are also performing much better, with positive average return on assets and cash return on assets; in contrast, the firms in the first decile of traditional operating accruals are among the worst performers in the sample with the lowest mean return on assets and cash return on assets. At the other extreme, the firm size and performance in the tenth decile are very similar in Panels A and B. As we show in the next table, this is because the firms overlap significantly at this end of the two different accrual sorts. Finally, Panel A shows that the percent operating accruals measure sorts on cash flows to some degree. Because Desai et al. (2004) document a trading strategy based on cash flows and contrast it with an accruals trading strategy, in the following sections we examine the overlap between the percent accruals measures and cash flow measures. The remaining descriptive statistics will be discussed in Section V as part of the exploration into why percent accruals are successful in identifying misvalued stocks.

Panels C and D of Table 2 give the distributions for the total accruals measures. Comparing the firms across the distributions of these two measures yields similar conclusions to the findings for the operating accrual measures, with one exception: sorting on percent total accruals does not sort on the cash return on assets.

Table 3 gives evidence on the overlap between a sort on percent accruals and a sort on traditional accruals or cash from operations. Generally, the overlap between percent operating accruals and traditional operating accruals is small in the low deciles but higher in the high deciles. Only 11.78 percent of the firms in the first decile of percent operating accruals are in the first decile of traditional operating accruals. This increases to 72.95 percent overlap in the tenth decile. A more general way to document the overlap between the two sorts is to give the average decile rank of one measure in each decile of the other measure. The decile rank of traditional operating accruals is 3.49 in the first decile of percent operating accruals (versus 1 if the two sorts were identical), and increases only marginally until the sixth decile where the average rank is 4.69. The average rank in the tenth decile is 9.70.

The sort on percent operating accruals produces a weak inverse sort on cash flows, as discussed earlier. The mean decile rank of cash from operations is 7.07 in the lowest percent operating accrual decile and is 2.44 in the highest decile (versus 10 and 1, respectively, if the two sorts were identical). The match between the two sorts is weak, however, with the average cash flow rank remaining above five until the ninth decile of percent operating accruals. In Section V we show that the returns to a percent operating accruals trading strategy are not driven by its correlation with cash from operations.

Table 3, Panel B shows the overlap between percent and traditional total accruals. For total accruals, the overlap is a bit stronger for low accruals and a bit weaker for high accruals. There is a 23.17 percent overlap in the first decile but only a 46.27 percent overlap in the tenth decile. The relevant cash flow comparison for total accruals is the net dividends/distributions plus the change in cash (i.e., total cash flows). Column three of Panel B shows that percent total accruals is picking out the extremes for total cash flows, with an average rank of 8.60 and 2.03 in the first and tenth deciles, respectively, although it does not appear to sort well through the middle of the distribution.

TABLE 2
Descriptive Statistics by Deciles of Various Accrual Measures

	1	2	3	4	5	6	7	8	9	10
Panel A: Deciles Sorted by Percent Operating Accrual										
Maximum Percent Operating Accrual	-4.04	-2.04	-1.34	-0.96	-0.67	-0.42	-0.18	0.14	0.82	
Mean Market Value	1,510	1,676	1,984	1,912	1,867	2,121	2,269	1,823	1,248	515
Mean Return on Assets	0.003	0.002	-0.014	-0.065	-0.091	-0.085	-0.093	-0.055	0.031	0.044
Mean Cash Return on Assets	0.118	0.117	0.108	0.084	0.051	0.017	-0.027	-0.034	-0.010	-0.079
Mean of Three-Year Sales Growth	0.082	0.091	0.084	0.088	0.103	0.119	0.130	0.135	0.129	0.116
Mean Book-to-Market	0.882	0.794	0.715	0.650	0.580	0.514	0.471	0.474	0.544	0.686
Percent of Firms with Positive Earnings	0.629	0.632	0.621	0.573	0.586	0.623	0.627	0.673	0.789	0.809
Percent of Firms with Special Item <0	0.228	0.212	0.227	0.267	0.262	0.194	0.157	0.114	0.089	0.090
Arbitrage Risk	0.139	0.138	0.135	0.148	0.156	0.155	0.159	0.159	0.158	0.170
Price per Share	13.92	15.66	17.41	16.70	17.38	18.93	19.01	18.18	16.83	12.03

	1	2	3	4	5	6	7	8	9	10
Panel B: Deciles Sorted by Traditional Operating Accrual										
Maximum Regular Operating Accrual	-0.20	-0.13	-0.09	-0.07	-0.05	-0.03	-0.02	0.01	0.06	
Mean Market Value	474	1,265	1,852	2,294	2,531	2,573	2,244	1,888	1,218	591
Mean Return on Assets	-0.383	-0.088	-0.024	0.004	0.014	0.023	0.024	0.032	0.031	0.047
Mean Cash Return on Assets	-0.032	0.070	0.085	0.085	0.075	0.066	0.050	0.037	0.002	-0.092
Mean of Three-Year Sales Growth	0.128	0.108	0.106	0.101	0.090	0.098	0.097	0.105	0.109	0.134
Mean Book-to-Market	0.501	0.663	0.667	0.672	0.662	0.661	0.654	0.656	0.616	0.557
Percent of Firms with Positive Earnings	0.132	0.415	0.595	0.691	0.744	0.787	0.796	0.800	0.799	0.801
Percent of Firms with Special Item <0	0.550	0.325	0.230	0.163	0.127	0.099	0.090	0.085	0.090	0.084
Arbitrage Risk	0.217	0.171	0.148	0.134	0.128	0.125	0.128	0.136	0.151	0.179
Price per Share	7.14	12.56	16.45	19.21	20.56	21.19	20.72	18.72	16.27	13.27

(continued on next page)

Panel C: Deciles Sorted by Percent Total Accrual

	1	2	3	4	5	6	7	8	9	10
Maximum Percent Total Accrual	-1.44	-0.89	-0.44	0.01	0.35	0.66	0.93	1.29	2.44	
Mean Market Value	1,247	770	1,067	1,946	2,839	3,103	2,369	1,446	1,092	779
Return on Assets	-0.023	-0.181	-0.201	-0.110	0.003	0.058	0.071	0.071	0.054	0.024
Cash Return on Assets	0.071	0.001	-0.040	-0.016	0.057	0.094	0.093	0.082	0.058	0.021
Total Cash Flow Return on Assets	0.099	0.024	-0.042	-0.060	-0.011	0.010	-0.002	-0.024	-0.076	-0.177
Mean Three-Year Sales Growth	0.072	0.074	0.116	0.096	0.102	0.109	0.111	0.107	0.111	0.108
Mean Book-to-Market	0.814	0.765	0.643	0.585	0.569	0.562	0.582	0.604	0.635	0.685
Percent of firms with positive earnings	0.462	0.172	0.294	0.563	0.787	0.894	0.940	0.939	0.891	0.777

Panel D: Deciles Sorted by Traditional Total Accrual

	1	2	3	4	5	6	7	8	9	10
Maximum Regular Total Accrual	-0.18	-0.07	-0.03	0.00	0.02	0.04	0.06	0.10	0.16	
Mean Market Value	402	919	1,465	2,148	2,361	2,534	2,342	2,008	1,533	959
Return on Assets	-0.399	-0.102	-0.023	0.008	0.025	0.037	0.048	0.059	0.070	0.050
Cash Return on Assets	-0.124	0.015	0.055	0.067	0.073	0.077	0.081	0.082	0.076	0.022
Total Cash Flow Return on Assets	-0.034	0.011	0.023	0.021	0.016	0.009	-0.001	-0.016	-0.054	-0.230
Mean Three-Year Sales Growth	0.124	0.084	0.077	0.076	0.081	0.085	0.095	0.110	0.126	0.148
Mean Book-to-Market	0.474	0.756	0.793	0.809	0.758	0.707	0.638	0.575	0.504	0.439
Percent of firms with positive earnings	0.096	0.269	0.453	0.686	0.844	0.897	0.909	0.900	0.889	0.782

The sample period is 1989–2008. Traditional operating accruals are defined as Net Income (Compustat item 172) less Cash from Operations (item 308) divided by average total assets (item 6). Traditional total accruals are computed as Net Income less Net Dividends/Distributions (item 115 + item 127 – item 108) and the change in cash (item 308 + item 311 + item 313) divided by average total assets. Percent Operating Accruals and Percent Total Accruals have the same numerators as Operating Accruals and Total Accruals, respectively, but the denominator for each is the absolute value of Net Income. Return on Assets is Net Income divided by average total assets. Cash Return on Assets is cash from operations divided by average total assets. Total Cash Flow Return on Assets is Net Dividends/Distributions plus the change in cash divided by average total assets. Special items (item 17) are coded as 0 if missing. Mean Three-Year Sales Growth is the average annual sales growth in the current year and prior two years. Market Value of Equity is measured at the fiscal year-end (item 25 times item 199). Price (item 199) and Book Value (item 60) are measured at the fiscal year-end. Arbitrage Risk is measured as the standard deviation of the residuals from a regression of firm-specific returns on the CRSP equally weighted index for up to 48 months preceding the month of return portfolio formation. All amounts are winsorized at 1 percent and 99 percent in this table and Table 1 only.

TABLE 3

Overlap between Percent Accruals, Traditional Accruals, and Cash Flow Deciles

Panel A: Percent Operating Accruals

Decile	Traditional Operating Accrual Decile Matches Percent Operating Accrual Decile	Decile Rank of Traditional Operating Accruals	Decile Rank of Cash from Operations Scaled by Total Assets
1	11.78%	3.49	7.07
2	20.38%	3.57	7.05
3	18.06%	3.66	6.76
4	14.75%	3.72	5.98
5	15.14%	3.99	5.54
6	21.14%	4.69	5.43
7	35.67%	5.67	5.24
8	58.33%	7.51	5.11
9	59.66%	9.03	4.40
10	72.95%	9.70	2.44

Panel B: Percent Total Accruals

Decile	Traditional Total Accrual Decile Matches Percent Total Accrual Decile	Decile Rank of Traditional Total Accruals	Decile Rank of Net Dividends and Change in Cash Scaled by Total Assets
1	23.17%	2.36	8.60
2	30.04%	2.01	6.23
3	29.23%	2.29	4.71
4	52.09%	3.57	5.82
5	49.58%	5.54	6.98
6	32.44%	6.73	7.09
7	25.15%	7.46	6.14
8	23.89%	7.91	4.45
9	26.65%	8.35	2.99
10	46.27%	8.80	2.03

Traditional operating accruals are defined as Net Income (Compustat item 172) less Cash from Operations (item 308) divided by average total assets (item 6). Traditional total accruals are computed as Net Income less Net Dividends/ Distributions (item 115 + item 127 - item 108) and the change in cash (item 308 + item 311 + item 313) divided by average total assets. Percent Operating Accruals and Percent Total Accruals have the same numerators as Operating Accruals and Total Accruals, respectively, but the denominator for each is the absolute value of Net Income.

To summarize, the lowest deciles of percent accruals are generally composed of different firms than the lowest deciles of traditional accruals—the firms for which the trading strategy will be long. The lowest decile of percent accruals firms are three times larger and much better performing than the lowest decile of traditional accruals, based on either operating or total accruals. The overlap between the percent accruals measures and the traditional accruals measures is much higher in the extreme positive deciles, although still far from perfect.

IV. FUTURE EXCESS RETURNS TO ACCRUAL TRADING STRATEGIES

In this section we document the return to the percent accruals trading strategy and compare it to the traditional accruals strategy and to a cash-from-operations strategy.

Excess Return Computation

Excess returns are computed as buy-and-hold size-adjusted returns, calculated as described in Barber et al. (1999). This method starts by constructing ten size portfolios based on December market value of equity of all NYSE firms. AMEX and NASDAQ firms are then placed in the appropriate size decile using the NYSE breakpoints. Because NYSE firms are typically much larger than AMEX and NASDAQ firms, this results in almost half the observations being sorted into the smallest portfolio. Barber et al. (1999) correct for this skewness by further subdividing the smallest portfolio into four size groups, so that in the end there are 14 size-referent portfolios.⁵ For each size portfolio and each month, the returns for each security in the portfolio are compounded over the following 12 months and then averaged across all the securities in the portfolio. The resulting annual return on each reference portfolio constitutes a passive, equally weighted investment in all securities in the portfolio. Excess returns are then the difference between the firm's annual return and the size-matched reference portfolio. For firms that are delisted during the future return period, we calculate the remaining return by taking CRSP's delisting return and then reinvesting the proceeds in the equally weighted reference portfolio. For firms delisted due to poor performance (delisting codes 500 and 520–584), we use a –35 percent delisting return for NYSE/AMEX firms and a –55 percent delisting return for NASDAQ firms, as recommended in Shumway (1997) and Shumway and Warther (1999). Note that the overall mean excess return is zero by construction. For all our results the return window is one year, beginning with the first day of the fourth month after the fiscal year-end.

All significance tests of excess returns are based on Fama-MacBeth t-statistics in order to control for the cross-sectional correlation between firms in a given decile for a given year. In particular, the 19 annual mean excess returns in a decile are averaged and the t-statistic is based on the standard deviation of these 19 observations.

We take extra care to ensure that the returns we report are the result of an implementable trading strategy. For each factor, we use the breakpoints between deciles from the previous year to divide the firms in the current year into ten portfolios. For example, for return windows starting in 1999, we use all observations between August 1998 (four months prior to January 1999) and September 1997 to construct a distribution for the factor and identify the ten breakpoints. This causes our ten portfolios to have slightly different numbers of observations, but makes it absolutely certain that the data necessary to sort the firms into portfolios are available when the return window opens.

Kraft et al. (2006) document a look-ahead bias in many accruals studies. Because many studies were interested in examining the evolution of accruals as well as the stock returns to an accruals-based trading strategy, the sample selection required that the next year's earnings be present. But whether earnings are present in the next year is not known at the time of portfolio formation, so the documented returns are not the result of an implementable trading strategy. Further, the firms with missing future data are not randomly selected; one reason data might be missing in the future is because the firm could have entered bankruptcy. Consistent with this conjecture, Kraft et al. (2006) show that, in a sample of NYSE/AMEX firms, the return to the low-accrual portfolio is 4.2 percent with the bias present but only 1.8 percent without it. In both Kraft et al. (2006) and our study, the excess return in the lowest traditional operating accrual decile is no longer significantly different from zero once the look-ahead bias is removed. The magnitude of the return in the lowest accrual decile is a particularly important part of the evidence supporting

⁵ Our results are very similar if we use only ten size portfolios. We prefer the 14-portfolio approach because it provides a closer size match for the smallest firms. Given that many market anomalies are more pronounced for small firms, controlling for size is particularly important.

the accruals anomaly. Without a significantly positive return to the long position in the lowest accrual decile, it is possible that the hedge return no longer exceeds transaction costs, especially given the high transaction costs associated with taking a short position in the high-accrual decile.

Return Results

Table 4 reports the returns by deciles sorted on percent and traditional operating accruals. The hedge return (decile 1 less decile 10) to percent operating accruals is 11.68 percent and is significant at $p < 0.001$. The hedge return comes equally from the long and short position, with an

TABLE 4
Mean Annual Size-Adjusted Returns to Percent Operating Accrual and Traditional Operating Accrual Portfolios

Panel A: Percent Operating Accruals

Decile	Maximum	Size-Adjusted Return	p-value	Number of Obs.
1	-4.04	0.0553	<0.001	8,210
2	-2.04	0.0516	<0.001	8,146
3	-1.34	0.0550	<0.001	8,139
4	-0.96	0.0416	<0.001	8,296
5	-0.67	0.0142	0.416	8,251
6	-0.42	0.0289	0.279	8,182
7	-0.18	0.0112	0.667	8,099
8	0.14	0.0029	>0.500	8,139
9	0.82	-0.0210	0.226	8,109
10		-0.0615	0.009	7,955
Decile 1 - Decile 10		0.1168	<0.001	

Panel B: Traditional Operating Accruals

Decile	Maximum	Size-Adjusted Return	p-value	Number of Obs.
1	-0.2023	0.0127	>0.500	8,237
2	-0.1286	0.0673	0.018	8,100
3	-0.0926	0.0444	0.092	8,197
4	-0.0696	0.0403	0.005	8,234
5	-0.0516	0.0266	0.024	8,040
6	-0.0343	0.0301	0.004	8,313
7	-0.0156	0.0167	0.087	8,155
8	0.0105	-0.0133	0.282	8,156
9	0.0594	-0.0106	0.490	8,127
10		-0.0524	0.029	7,973
Decile 1 - Decile 10		0.0651	0.186	

Returns are the time-series mean annual buy-and-hold size-adjusted returns, calculated as described in Barber et al. (1999) beginning in the fourth month after the fiscal year-end. The p-values are based on two-tailed Fama-MacBeth t-statistics computed over the 19 annual mean returns in a decile and the standard deviation of these 19 observations. The breakpoints between deciles are based on the previous year's cutoffs, so the number of observations in each decile varies slightly. Traditional operating accruals are defined as Net Income (Compustat item 172) less Cash from Operations (item 308) divided by average total assets (item 6). Percent operating accruals have the same numerator but scale by the absolute value of net income.

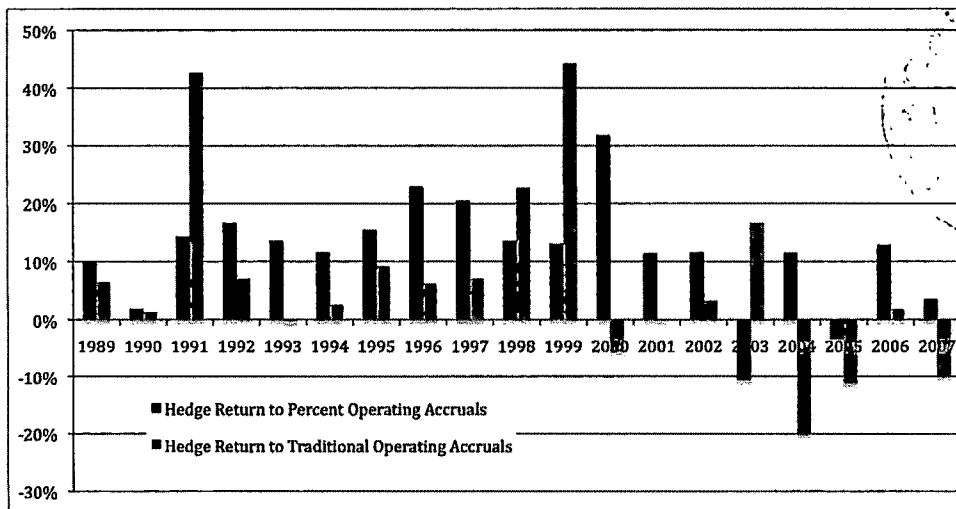
excess return of 5.53 percent in decile 1 and an excess return of -6.15 percent in decile 10; both are significant. In contrast, the hedge return to traditional operating accruals in Panel B is 6.51 percent, but is insignificant with a p -value of 0.186. Further, the excess return in decile 1 is only 1.27 percent and is not significant. Consistent with the fact that the overlap between the two accrual definitions is high for the top decile, the excess returns in decile 10 of traditional operating accruals is -5.24 percent and is significant ($p = 0.029$).⁶ Figure 2 compares the hedge returns to percent and traditional operating accrual strategies for each of the 19 years in the sample. The hedge based on percent operating accruals is larger in 15 of the 19 years; a simple binomial test is significant ($p < 0.001$).

Table 5 reports the returns to percent total accruals and traditional total accruals. Panel A shows that the hedge return to percent total accruals is 8.53 percent and significant ($p = 0.001$). The long position in decile 1 contributes 5.49 percent of the hedge return, and is significant ($p = 0.005$). In contrast, Panel B shows that the hedge return to traditional total accruals is not significant. The excess return to the long position is only 1.60 percent and insignificant, although the excess return to decile 10 is -4.23 percent and significant ($p = 0.048$).

In summary, the percent accruals definitions appear to dominate the traditional accrual definitions for both operating and total accruals. Percent accruals generate larger hedge returns, larger excess returns in decile 1 where the two definitions differ the most, and more significant results in all the extreme deciles.

The fact that the accrual anomaly is large and significant when measured as percent accruals offers researchers an empirical regularity that is difficult to reconcile with an efficient stock

FIGURE 2
Annual Size-Adjusted Hedge Returns to Percent and Traditional Operating Accruals (by year in which the 12-month return window begins)



⁶ There are 967 observations that are in both the first decile of percent operating accruals and the first decile of traditional operating accruals. The mean return for these firms is 7.1 percent, although it is not statistically significant ($p = 0.110$).

TABLE 5
Mean Annual Size-Adjusted Returns to Percent Total Accrual and Traditional Total Accrual Portfolios

Panel A: Percent Total Accruals

Decile	Maximum	Size-Adjusted Return	p-value	Number of Obs.
1	-1.436	0.0549	0.005	7,461
2	-0.890	0.0383	0.110	7,400
3	-0.439	0.0162	>0.500	7,290
4	0.008	0.0387	0.242	7,357
5	0.350	0.0178	0.240	7,193
6	0.658	0.0239	0.035	7,290
7	0.929	0.0076	>0.500	7,225
8	1.291	0.0165	0.332	7,389
9	2.437	-0.0093	0.289	7,264
10		-0.0304	0.092	7,264
Decile 1 – Decile 10		0.0853	0.001	

Panel B: Traditional Total Accruals

Decile	Maximum	Size-Adjusted Return	p-value	Number of Obs.
1	-0.177	0.0160	>0.500	7,418
2	-0.073	0.0604	0.133	7,446
3	-0.027	0.0325	0.028	7,324
4	0.000	0.0420	0.002	7,270
5	0.019	0.0139	0.429	7,227
6	0.037	0.0012	>0.500	7,237
7	0.060	0.0212	0.236	7,272
8	0.095	0.0357	0.005	7,306
9	0.165	-0.0009	>0.500	7,341
10		-0.0423	0.048	7,298
Decile 1 – Decile 10		0.0583	0.313	

Returns are the time-series mean annual buy-and-hold size-adjusted returns, calculated as described in Barber et al. (1999) beginning in the fourth month after the fiscal year-end. The p-values are based on two-tailed Fama-MacBeth t-statistics computed over the 19 annual mean returns in a decile and the standard deviation of these 19 observations. The breakpoints between deciles are based on the previous year's cutoffs, so the number of observations in each decile varies slightly. Traditional total accruals are computed as Net Income less Net Dividends/Distributions (item 115 + item 127 – item 108) and the change in cash (item 308 + item 311 + item 313) divided by average total assets, defined as Net Income (Compustat item 172) less Cash from Operations (item 308) divided by average total assets (item 6). Percent total accruals have the same numerator but scale by the absolute value of net income.

market. It is also inconsistent with the Kothari et al. (2006) agency hypothesis, insofar as the long position contributes a large and statistically significant portion of the excess return. In the next section we explore why extreme values of percent accruals identify mispriced stocks.

V. WHAT DRIVES THE RETURNS TO PERCENT ACCRUALS?

In this section, we explore the source of the excess returns in the portfolios of extreme percent accruals, focusing in particular on the long position in the lowest decile. For brevity, we only consider percent operating accruals, but the results and conclusions are similar for percent total

accruals. Recall that only 12 percent of the observations in the first decile of the traditional operating accruals measure are in the first decile of percent operating accruals. What is the difference between the newly constructed accrual portfolios and the traditional ones, particularly in the lowest accruals portfolio? We offer one explanation and then eliminate a number of other tempting-but-wrong explanations for our return results.

One reason percent operating accruals identifies misvalued stocks is that this measure successfully sorts observations that are more “extreme” into the extreme portfolios than does the traditional operating accruals measure. To understand this, consider the naïve and sophisticated forecasting models posited in the “earnings fixation” hypothesis. Because the scaling variable is what is at issue, we specify each model with unscaled variables and then control for size by including common equity (item 60) as another independent variable, as discussed in Barth and Kallapur (1996), and we estimate the regression on percentile ranks.⁷ The naïve model is:

$$\text{Next Year's Net Income} = \alpha_0 + \alpha_1 * \text{Net Income} + \alpha_2 * \text{Common Equity} + \varepsilon.$$

The more sophisticated model distinguishes between cash flows and accruals, resulting in:

$$\begin{aligned} \text{Next Year's Net Income} = & \beta_0 + \beta_1 * \text{Cash from Operations} + \beta_2 * \text{Operating Accruals} \\ & + \beta_3 * \text{Common Equity} + \varepsilon. \end{aligned}$$

Noting that net income is identically equal to cash from operations plus operating accruals, the expected difference between the predictions of the two models is:

$$\begin{aligned} \text{Difference} = & (\beta_0 - \alpha_0) + (\beta_1 - \alpha_1) * \text{Cash from Operations} + (\beta_2 - \alpha_1) * \text{Operating Accruals} \\ & + (\beta_3 - \alpha_2) * \text{Common Equity}. \end{aligned}$$

When the difference is positive, the sophisticated model predicts higher next year's net income than the naïve model. Under the earnings fixation hypothesis, this approach identifies a stock as being undervalued—the market behaves as if it is using the naïve model and is subsequently surprised when next year's net income is higher than expected. Consistent with Sloan (1996) and a host of other studies, we show later that $\beta_1 > \alpha_1 > \beta_2$; that is, cash from operations is more persistent than accruals, and the persistence of earnings lies somewhere in between. With this, the most extreme positive differences occur when cash from operations is an extreme positive amount *and* operating accruals is an extreme negative amount. These are exactly the observations that percent operating accruals sorts into the first decile. The large negative accrual results in a large negative numerator and the large positive cash flow of approximately the same magnitude makes the denominator small. In contrast, while the first decile of the traditional accruals contains large negative accruals, it also contains many firms with large negative cash from operations (as seen in Table 2).

To quantify this effect, Table 6, Panel A reports the coefficients from the naïve and sophisticated models, estimating each regression using the percentile ranks of the variables (as in Sloan 1996). To partially control for the look-ahead bias inherent in this type of analysis, if the firm is missing next year's earnings, then it is set equal to the negative of this year's common equity. By writing off all the book value, we are effectively assuming the firm was liquidated. The results in

⁷ Whether we control for size by including Common Equity in the model has little effect. The coefficient on this variable is almost the same in the naïve and sophisticated models, and so its effect washes out when computing the difference between the models. Consequently, the results in Table 6 are almost identical without this variable.

TABLE 6

Difference between Sophisticated and Naïve Forecasts in Different Accrual Portfolios

Panel A: Percentile Rank Regression of Next Year's Earnings (unscaled)

	Intercept	Current Earnings	Cash from Operations	Operating Accruals	Common Equity
Naïve model	21.96	0.505			0.051
p-value	<0.001	<0.001			<0.001
Sophisticated model	8.86		0.575	0.198	0.048
p-value	<0.001		<0.001	<0.001	<0.001

Panel B: Mean Differences between Sophisticated and Naïve Forecasts by Accrual Decile

Decile	Sorted by Traditional Operating Accruals	Sorted by Percent Operating Accruals
1*	4.6	8.6
2	4.6	5.9
3	2.7	4.3
4	1.3	1.3
5	0.0	-2.4
6	-1.1	-3.6
7	-1.6	-3.1
8	-0.7	-0.8
9	-1.9	-0.8
10*	-8.4	-9.9

* Significantly different using either a two-sample t-test or a Wilcoxon test at the <0.001 level.

For each observation we compute the naïve forecast and the sophisticated forecast based on the models in Panel A, and then compute the difference in forecasts. Panel B reports the mean difference in forecasts between the two models for each decile of accruals, sorted on either traditional operating accruals or percent operating accruals. Earnings is data item 172, cash from operations is data item 308, and operating accruals is earnings less cash from operations. Common Equity is data item 60. The dependent variable is next year's earnings. All variables are unscaled. Missing values of the dependent variable are replaced with the negative value of current common equity (item 60).

Table 6 are not sensitive to this assumption—simply eliminating the missing observations yields similar results. As evident in the table, all coefficients are highly significant, and the relative persistence is consistent with prior studies; cash flows are the most persistent, followed by earnings, then followed by accruals.

In Table 6, Panel B we quantify exactly how sorting by percent accruals selects more extreme differences between the naïve and sophisticated models. For each observation, we compute the naïve forecast, the sophisticated forecast, and the difference between the two forecasts. We then report the mean of these differences sorted by percent operating accruals and traditional operating accruals. For instance, in decile 1 of traditional operating accruals, the sophisticated forecast is, on average, 4.6 percent larger than the naïve forecast. By comparison, the mean difference between the forecasts in decile 1 of percent operating accruals is 8.6 percent, almost twice as extreme as the difference for traditional operating accruals. Similarly, the difference in forecasts is -8.4 percent in the tenth decile of traditional operating accruals, as compared to -9.9 percent for the tenth decile of percent operating accruals. The differences in the first decile and tenth decile portfolios are significant at the 0.001 level.

Bradshaw et al. (2001) provide evidence that analysts fail to fully account for the information in accruals in their forecasts. If extreme percent operating accruals are better than traditional accruals at identifying observations where there are extreme differences between the naïve and sophisticated forecasts, then it is possible that they will also do a better job better describing the error in analyst forecasts. To examine this possibility, we regress signed analyst forecast errors of earnings per share on either traditional operating accruals or percent operating accruals. The forecast is the consensus median forecast in the fourth month following the fiscal year-end (the same month that the return window opens) and the realization is for the following fiscal year, both taken from I/B/E/S. Table 7 reports the results. The accruals measures are converted to their decile rank and then scaled to fall in $[0, 1]$. Consistent with Bradshaw et al. (2001), the coefficient on traditional operating accruals is -0.0039 with a t -statistic of -6.1 . Analysts systematically overestimate earnings per share for high-accrual firms and underestimate earnings per share for low-accrual firms. But more importantly, the relation is stronger when percent operating accruals are used, with a coefficient of -0.0049 and a t -statistic of -8.2 .

In sum, by conditioning the accrual on the level of net income, percent operating accruals effectively pick out more extreme combinations of cash flows and accruals, resulting in more extreme differences in the naïve and sophisticated forecast models. Consistent with this pattern, analyst forecast errors have a stronger negative association with percent operating accruals than with traditional operating accruals. This evidence supports the earnings fixation hypothesis, insofar as the mispricing of accruals is greatest in places where the hypothesis predicts the greatest differences in naïve and sophisticated beliefs.

Eliminating Some Alternative Explanations for the Success of Percent Accruals

In this section, we run through a number of other candidate reasons for the superior performance of the percent operating accruals.

TABLE 7
Regression of Analyst Forecast Errors on Traditional Operating Accruals and Percent Operating Accruals

	Intercept	Decile Rank of Traditional Operating Accrual	Decile Rank of Percent Operating Accrual
Model 1	-0.0107	-0.0039	
t-statistic	-24.2	-6.1	
p-value	<0.001	<0.001	
Model 2	-0.0102		-0.0049
t-statistic	-25.0		-8.2
p-value	<0.001		<0.001

The forecast error is measured as the actual earnings per share less the median forecast in the fourth month following the fiscal year-end (i.e., the first forecast after the portfolio formation month), taken from I/B/E/S. Traditional operating accruals are defined as Net Income (Compustat item 172) less Cash from Operations (item 308) divided by average total assets (item 6). Percent operating accruals have the same numerator but scale by the absolute value of net income. The independent variables are ranked into deciles, then scaled to be in $[0, 1]$. The sample size is 40,352.

Percent Accruals Proxies for Cash from Operations

Desai et al. (2004) document a powerful trading strategy based on the ratio of cash from operations to price, and argue that accruals are simply a proxy for this variable, given the strong negative correlation between accruals and cash from operations. They scale by price to capture the well-known value-glamour anomaly (Lakonishok et al. 1994). The value-glamour effect is outside our scope, but in Table 9 of their article, they regress future returns on cash from operations scaled by total assets and accruals scaled by total assets. They find that accruals are no longer significantly related to future returns, but cash from operations remains highly significant. It is therefore possible that percent operating accruals proxy for a more fundamental cash-from-operations anomaly. To examine this possibility, we first document the hedge return to cash from operations scaled by total assets, shown in Table 8. We do not find a significant hedge return, but the return to the long position in the high cash-from-operations portfolio is 6.62 percent and significant at less than 0.001 level. In Table 9, we estimate regressions of next year's returns on the decile rank of percent operating accruals and cash from operations, where the rank is scaled to fall in $[0, 1]$. In model 1, the coefficient on percent operating accruals is -0.0993 , implying that the expected hedge between the extreme deciles is 9.93 percent, and is significant ($p = 0.001$). In contrast, the coefficient on cash from operations is insignificant in model 2. And finally, when both variables are included in the regression in model 3, the estimated return and significance of percent operating accruals falls slightly, but is still significant, while cash from operations remains insignificant. In sum, sorting on percent operating accruals weakly sorts on cash from operations, as shown in Table 2, but the association with cash flows is not the driving force behind the success of percent operating accruals.

TABLE 8
Mean Annual Size-Adjusted Returns to Cash from Operations Scaled by Total Assets Portfolios

Decile	Maximum	Size-Adjusted Return	p-value	Number of Obs.
1	-0.1522	-0.0496	0.474	8,240
2	-0.0362	0.0030	>0.500	8,043
3	0.0139	0.0026	>0.500	8,041
4	0.0442	0.0145	0.443	8,040
5	0.0681	0.0273	0.049	8,276
6	0.0900	0.0273	0.024	8,092
7	0.1140	0.0378	0.003	8,150
8	0.1435	0.0350	0.009	8,132
9	0.1919	0.0307	0.030	8,251
10		0.0662	<0.001	8,267
Decile 10 – Decile 1		0.1158	0.113	

Returns are the time-series mean annual buy-and-hold size-adjusted returns, calculated as described in Barber et al. (1999) beginning in the fourth month after the fiscal year-end. The p-values are based on two-tailed Fama-MacBeth t-statistics computed over the 19 annual mean returns in a decile and the standard deviation of these 19 observations. The breakpoints between deciles are based on the previous year's cutoffs, so the number of observations in each decile varies slightly. Portfolios are sorted by Cash from Operations (item 308) divided by average total assets (item 6).

TABLE 9
Regressions of Annual Size-Adjusted Returns on Percent Operating Accruals and Cash from Operations

	Intercept	Decile Rank of Percent Operating Accrual	Decile Rank of Cash from Operations Scaled by Total Assets
Model 1	0.0674	-0.0993	
p-value	<0.001	0.001	
Model 2	-0.0212		0.0807
p-value	0.624		0.218
Model 3	0.0346	-0.0784	0.0474
p-value	0.483	<0.001	0.485

Returns are the time-series mean annual buy-and-hold size-adjusted returns, calculated as described in Barber et al. (1999) beginning in the fourth month after the fiscal year-end. The p-values are based on two-tailed Fama-MacBeth t-statistics computed over the 19 annual mean returns in a decile and the standard deviation of these 19 observations. The breakpoints between deciles are based on the previous year's cutoffs, so the number of observations in each decile varies slightly. Percent operating accruals are defined as Net Income (Compustat item 172) less Cash from Operations (item 308), divided by the absolute value of Net Income. The independent variables are ranked into deciles then scaled to be in [0, 1].

Special Items

A common reason a firm has large, negative, unscaled accruals is because it records a non-cash special item, such as an asset write-off, in the period. Dechow and Ge (2006) study this explanation for the traditional accruals measure, reporting that the positive returns in the low-accrual decile in their sample is driven by firms with large negative special items. Indeed, Table 2, Panel B shows that 55.00 percent of the firms in the first decile of traditional accruals have negative special items. In contrast, only 22.83 percent of the firms in the lowest decile of percent accruals have a negative special item. In Table 10, we partition each decile into observations with or without a negative special and examine the returns within each partition.⁸

Table 10, Panel A shows that for both subsamples the hedge return to percent accruals is large and significant; it is 11.88 percent when there is no negative special item and 9.94 percent when a special item is present. Further, the excess return in the first decile is approximately the same size regardless of the special item. Table 10, Panel B shows results consistent with the unpartitioned results for traditional operating accruals; the hedge return is insignificant in both subsamples, as are the returns in the first decile. In sum, the returns to a trading strategy based on percent operating accruals are not particularly sensitive to the presence or absence of a negative special item.

Gains versus Losses

Table 11 bisects the data based on whether the firm reported positive or negative net income. When the data are sorted based on traditional operating accruals, the results are very similar to

⁸ We consider two alternative partitions of the sample: (1) the definition in Dechow and Ge (2006) requires that the negative special item exceeds 2 percent of total assets, selecting about 20 percent of the sample, and (2) a definition in the spirit of percent accruals is that the negative special item exceed 25 percent of absolute net income, which also selects about 20 percent of the sample. We prefer using the (unscaled) existence of a negative special item because it yields the same partition regardless of the scale variable used. The conclusions in Table 10 are qualitatively similar using either of these alternative definitions.

TABLE 10

Mean Annual Size-Adjusted Returns for Operating Accruals by Special Item Subsamples

Panel A: Percent Operating Accruals

No Negative Special Item (64%)

Special Item < 0 (36%)

Decile	Return	p-value	% of sample	Return	p-value	% of sample
1	0.0511	0.003	5.5%	0.0604	0.002	4.5%
2	0.0414	0.004	5.7%	0.0698	0.002	4.3%
3	0.0449	0.005	5.7%	0.0745	0.001	4.3%
4	0.0446	0.007	5.6%	0.0355	0.076	4.5%
5	0.0112	0.375	5.8%	0.0020	>0.500	4.3%
6	0.0021	>0.500	6.5%	0.0532	0.354	3.5%
7	0.0064	>0.500	6.9%	0.0194	>0.500	3.1%
8	-0.0012	>0.500	7.4%	-0.0116	>0.500	2.6%
9	-0.0272	0.077	7.6%	0.0078	>0.500	2.4%
10	-0.0678	0.019	7.4%	-0.0390	0.171	2.3%
D1 - D10	0.1188	<0.001		0.0994	0.035	

Panel B: Traditional Operating Accrual

No Negative Special Item (64%)

Special Item < 0 (36%)

Decile	Return	p-value	% of sample	Return	p-value	% of sample
1	-0.0446	0.125	3.8%	0.0416	0.466	6.3%
2	0.0581	0.006	5.2%	0.0742	0.053	4.7%
3	0.0331	0.153	6.0%	0.0675	0.082	4.1%
4	0.0349	0.000	6.4%	0.0553	0.165	3.7%
5	0.0260	0.023	6.5%	0.0238	>0.500	3.4%
6	0.0244	0.028	6.9%	0.0890	0.042	3.2%
7	0.0158	0.123	7.0%	0.0183	>0.500	3.0%
8	-0.0134	0.301	7.3%	0.0192	>0.500	2.7%
9	-0.0095	>0.500	7.4%	-0.0369	>0.500	2.6%
10	-0.0569	0.028	7.7%	0.0254	>0.500	2.1%
D1 - D10	0.0123	>0.500		0.0162	0.197	

Returns are the time-series mean annual buy-and-hold size-adjusted returns, calculated as described in Barber et al. (1999) beginning in the fourth month after the fiscal year-end. The p-values are based on two-tailed Fama-MacBeth t-statistics computed over the 19 annual mean returns in a decile and the standard deviation of these 19 observations. The breakpoints between deciles are based on the previous year's cutoffs, so the number of observations in each decile varies slightly. Traditional operating accruals are defined as Net Income (Compustat item 172) less Cash from Operations (item 308) divided by average total assets (item 6). Percent operating accruals have the same numerator but scale by the absolute value of net income. Special Items (item 17) are coded as 0 if missing.

those reported by Dopuch et al. (2009), with no significant returns for any decile in the loss subsample, as shown in Panel B. This is a particularly troublesome result for the traditional accrual anomaly, as 34 percent of the observations in the population have losses, and there is nothing in the naïve investor hypothesis that says the result should only hold for firms with positive net income. However, when the data are sorted based on percent operating accruals, Panel A of Table 11 shows that the hedge return and the excess return for the lowest decile are significant in both subsamples, and actually somewhat larger and more significant in the loss subsample than in the gain subsample.

TABLE 11

Mean Annual Size-Adjusted Returns for Operating Accruals by Gain/Loss Subsamples

Panel A: Percent Operating Accruals

Decile	Loss (34%)		% of Sample	Gain (66%)		% of Sample
	Return	p-value		Return	p-value	
1	0.0715	<0.001	3.7%	0.0459	<0.001	6.4%
2	0.0487	0.032	3.7%	0.0533	0.001	6.3%
3	0.0712	0.031	3.8%	0.0408	0.042	6.2%
4	0.0435	0.131	4.4%	0.0379	0.024	5.9%
5	-0.0146	>0.500	4.2%	0.0204	0.224	5.9%
6	0.0143	>0.500	3.8%	0.0233	0.110	6.3%
7	-0.0080	>0.500	3.7%	0.0210	0.075	6.3%
8	-0.0145	>0.500	3.3%	0.0105	>0.500	6.7%
9	-0.0740	0.101	2.1%	-0.0193	0.146	7.9%
10	-0.0636	0.155	1.9%	-0.0634	0.002	7.9%
D1 - D10	0.1351	0.007		0.1093	<0.001	

Panel B: Traditional Operating Accruals

Decile	Loss (34%)		% of sample	Gain (66%)		% of sample
	Return	p-value		Return	p-value	
1	-0.0021	>0.500	8.8%	0.0619	0.128	1.3%
2	0.0539	0.195	5.8%	0.0743	<0.001	4.1%
3	0.0253	>0.500	4.1%	0.0482	0.025	6.0%
4	0.0350	0.361	3.1%	0.0374	0.002	7.0%
5	0.0075	>0.500	2.5%	0.0290	0.137	7.3%
6	0.0506	0.258	2.2%	0.0237	0.132	8.0%
7	0.0357	0.458	2.0%	0.0116	0.411	8.0%
8	-0.0414	0.259	2.0%	-0.0061	>0.500	8.0%
9	-0.0359	0.496	2.0%	-0.0069	>0.500	8.0%
10	-0.0468	0.384	1.9%	-0.0578	0.004	7.8%
D1 - D10	0.0447	>0.500		0.1197	0.009	

Returns are the time-series mean annual buy-and-hold size-adjusted returns, calculated as described in Barber et al. (1999) beginning in the fourth month after the fiscal year-end. The p-values are based on two-tailed Fama-MacBeth t-statistics computed over the 19 annual mean returns in a decile and the standard deviation of these 19 observations. The breakpoints between deciles are based on the previous year's cutoffs, so the number of observations in each decile varies slightly. Traditional operating accruals are defined as Net Income (Compustat item 172) less Cash from Operations (item 308) divided by average total assets (item 6). Percent operating accruals have the same numerator but scale by the absolute value of net income. Loss or Gain is based on the sign of item 172.

Partitioning the percent accruals strategy on gain versus loss offers an interesting connection to Burgstahler and Dichev (1997). In this study, firms with small gains are considered more likely to be manipulating earnings up than firms with small losses. As previously discussed, the extreme deciles of percent accruals are firms with relatively small positive or small negative earnings (making the denominator close to zero). Consistent with the earnings manipulation hypothesis, Table 11, Panel A shows that firms that are in the first decile of percent accruals, and accrue down into a loss position, have higher returns in the subsequent year, as compared to those that accrued down, but into a small gain position. The argument is that their accruals are less suspicious

because they chose to accrue into a small loss. At the other extreme, the earnings manipulation hypothesis suggests that firms in the top decile of percent accruals that accrue up to a small gain are more “manipulative” than those that accrue up to a small loss. The results for this case are not as clear; the returns in the top decile of percent accruals are very similar regardless of the sign of the firm’s earnings. However, the significance level of the negative subsequent returns is superior for firms that accrued up to a small positive gain, as compared to those that accrued up to a small loss. Without overstating the significance of these results, there is some evidence that extreme percent accruals combined with the sign of earnings might flag firms that have manipulated earnings in a way that results in predictable market returns in the subsequent year.

Limits to Arbitrage

Mashruwala et al. (2006) find that the accrual anomaly, based on traditional operating accruals, is concentrated in firms with high arbitrage risk, as measured by the idiosyncratic volatility of their stock returns. They also find that it is concentrated in firms with lower share prices, which is a common proxy for transaction costs. They suggest that the combination of risk imposed on arbitrageurs and high transaction costs might leave the accrual anomaly difficult to exploit. To compare the two measures of accruals’ exposure to arbitrage risk, we examine the returns to each strategy for each quintile of arbitrage risk. Following Mashruwala et al. (2006), we estimate arbitrage risk as the standard deviation of residuals from a regression of the firm-level returns on the CRSP equally weighted index for up to 48 months preceding the month of return portfolio formation.⁹ To keep the portfolio sizes roughly comparable to previous results, we sort percent accruals and traditional accruals into five quintiles. The resulting two-way sort is given in Table 12, and shown graphically in Figure 3.

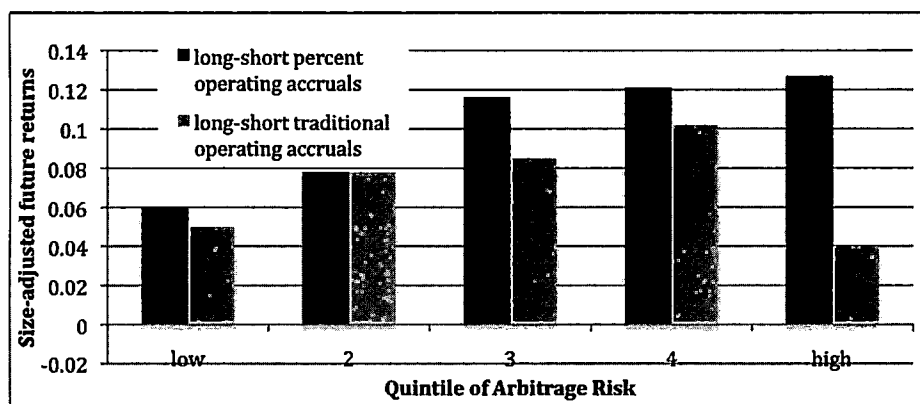
As seen in Table 12, Panel A, for every quintile of arbitrage risk, the hedge return to percent operating accruals dominates the return to traditional operating accruals. More importantly, the return to percent accruals exceeds 10 percent in the top three quintiles, or over half the sample, and is significant in all but the lowest quintile. Figure 3, Panel B shows the size-adjusted returns to the long position of each strategy across quintiles of arbitrage risk. The returns to the long position for both strategies is essentially zero for the quintile with the lowest arbitrage risk, but for the top three quintiles the percent accruals strategy dominates the traditional accrual strategy, and yields size-adjusted returns of 7 percent, 11 percent, and 8 percent. Table 12, Panel A shows that the long-only returns are significant in the top three quintiles of arbitrage risk, and the hedge returns are significant in the top four quintiles of arbitrage risk. In sum, the results in Figure 3 and Table 12 show that, while percent accruals are not completely insensitive to the level of arbitrage risk, they continue to be large and significant for a much larger portion of the distribution of arbitrage risk than traditional accruals.

With respect to transaction costs, Panels A and B of Table 2 shows that the firms in the lowest decile of percent operating accruals are considerably larger than the firms in the lowest decile of regular operating accruals. They also have considerably higher share prices. The mean share price in decile 1 for traditional operating accruals is \$7.14 per share, while it is \$13.92 per share for percent operating accruals, almost twice as large. Larger firms and higher share prices have been shown to proxy for lower transaction costs (e.g., Bartov et al. 2000). It is therefore unlikely that greater transaction costs are allowing the higher excess returns to the first decile of percent operating accruals to remain unexploited.

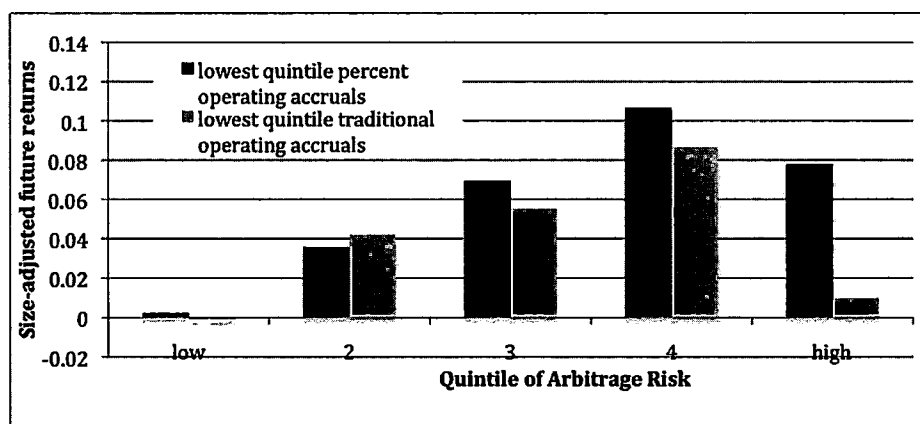
⁹ Mashruwala et al. (2006) require that the firm have 48 prior months of returns to be included in the sample. We do not impose this data requirement because it is not generally a requirement imposed on studies of the accrual anomaly. Further, as they report, the returns to an accrual strategy are significantly greater when this data requirement is not imposed. Finally, while estimates based on fewer than 48 observations will be noisier, they will still be unbiased. Requiring 48 months of prior returns, on the other hand, could inadvertently introduce a bias.

FIGURE 3
Annual Size-Adjusted Returns to Percent Operating Accruals by Quintile of Arbitrage Risk

Panel A: Return to Hedge between Top and Bottom Percent Operating Accrual Quintiles



Panel B: Return to Long Position in Lowest Percent Operating Accrual Quintile



Some other candidate explanations for the success of percent operating accruals are as follows:

Greater Association with Growth? The association between accrual measures and “growth” depends on how growth is measured; if it is measured as the increase in operating assets, for instance, then operating accruals and growth are almost synonymous. To assess whether percent accruals are proxying for an underlying “growth” anomaly, without imposing a mechanical relation between the two measures, we compute the average annual sales growth rate over the past three years, ending in the year in which accruals are measured. We use this definition of growth because it is not directly based on changes in operating assets in the same

TABLE 12
Mean Annual Size-Adjusted Returns to Percent and Traditional Accrual Portfolios by Arbitrage Risk Quintiles

Panel A: Percent Operating Accruals

		Quintile of Arbitrage Risk				
		1	2	3	4	5
1	return	0.0024	0.0356	0.0695	0.1056	0.0779
	p-value	>0.500	0.120	0.011	<0.001	0.076
	# of Obs.	3,307	4,114	3,784	3,031	2,126
	return	0.0094	0.0308	0.0427	0.0774	0.0817
	p-value	>0.500	0.075	0.012	0.029	0.140
2	# of Obs.	4,056	3,371	3,267	3,068	2,673
	return	-0.002	0.0259	0.0384	0.053	-0.017
	p-value	>0.500	0.115	0.139	0.250	>0.500
	# of Obs.	3,914	2,881	2,811	2,995	3,832
	return	-0.01	-0.003	0.0383	0.0211	-0.005
3	p-value	>0.500	>0.500	0.064	>0.500	>0.500
	# of Obs.	3,298	2,967	2,852	3,279	3,842
	return	-0.058	-0.042	-0.046	-0.015	-0.049
	p-value	0.041	0.136	0.030	>0.500	0.235
	# of Obs.	1,724	2,979	3,598	3,933	3,830
4	return	0.0604	0.0776	0.1155	0.1206	0.1269
	p-value	0.106	0.032	<0.001	0.016	0.034
	# of Obs.	5,031	7,093	7,382	6,964	5,956
5	return	0.0024	0.0356	0.0695	0.1056	0.0779
	p-value	>0.500	0.120	0.011	<0.001	0.076
	# of Obs.	3,307	4,114	3,784	3,031	2,126
	return	0.0094	0.0308	0.0427	0.0774	0.0817
	p-value	>0.500	0.075	0.012	0.029	0.140
Q1 - Q5	# of Obs.	4,056	3,371	3,267	3,068	2,673
	return	-0.002	0.0259	0.0384	0.053	-0.017
	p-value	>0.500	0.115	0.139	0.250	>0.500
	# of Obs.	3,914	2,881	2,811	2,995	3,832
	return	-0.01	-0.003	0.0383	0.0211	-0.005

(continued on next page)

Panel B: Traditional Operating Accruals

		Quintile of Arbitrage Risk				
		1	2	3	4	5
1	return	-0.0010	0.0420	0.0553	0.0865	0.0099
	p-value	>0.500	0.039	0.018	0.065	>0.500
	# of Obs.	1,120	2,157	2,986	4,099	5,975
	return	0.0176	0.0335	0.0601	0.0678	0.0310
	p-value	0.459	0.038	0.006	0.130	>0.500
2	# of Obs.	3,590	3,823	3,532	3,047	2,439
	return	0.0009	0.0283	0.0374	0.0608	0.0535
	p-value	>0.500	0.135	0.063	0.110	0.218
	# of Obs.	5,354	3,675	2,968	2,444	1,912
	return	-0.0200	-0.0110	0.0280	0.0265	-0.0100
3	p-value	0.461	>0.500	0.074	>0.500	>0.500
	# of Obs.	4,481	3,711	3,251	2,785	2,083
	return	-0.0510	-0.0360	-0.0290	-0.0150	-0.0300
	p-value	0.056	0.182	0.194	>0.500	>0.500
	# of Obs.	1,754	2,946	3,575	3,931	3,894
4	return	0.0500	0.0780	0.0843	0.1015	0.0399
	p-value	0.213	0.021	0.009	0.076	>0.500
	# of Obs.	2,874	5,103	6,561	8,030	9,869
5	return					
	p-value					
	# of Obs.					
	return					
	p-value					
Q1 - Q5	# of Obs.					
	return					
	p-value					
	# of Obs.					
	return					

Returns are the time-series mean annual buy-and-hold size-adjusted returns, calculated as described in Barber et al. (1999) beginning in the fourth month after the fiscal year-end. The p-values are based on two-tailed Fama-MacBeth t-statistics computed over the 19 annual mean returns in a quintile. Arbitrage Risk is measured as the standard deviation of the residuals from a regression of firm-specific returns on the CRSP equally weighted index for up to 48 months preceding the month of return portfolio formation.

period, and because it has been used in prior studies as a measure of “growth” or “glamour” (see Desai et al. 2004, for example). Table 2, Panel A shows how growth varies across percent operating accrual deciles. The table shows only a mild increase in growth across deciles. In the extreme deciles, growth is lower in the first and tenth deciles of percent operating accruals than it is in the first and tenth deciles of traditional operating accruals. As an alternative growth measure, we compute that current year’s sales growth. For our sample, the Spearman correlation between the current year’s sales growth and either traditional or percent operating accruals is 0.19 and, like the past three-year growth rate, is lower in the extreme deciles of percent accruals than in the extreme deciles of traditional accruals. In sum, while percent operating accruals is weakly correlated with sales growth, it is no more correlated than traditional operating accruals.

Faster Mean Reversion in Accruals? Earlier we reported that the differences between the naïve and sophisticated forecasting models are larger in the extreme deciles of percent operating accruals than in the extreme deciles of traditional operating accruals. This pattern does not lead to a more rapid mean reversion in accruals, however. Forty-one percent of the firms in the first two deciles of traditional operating accruals remain in the first two deciles in the following year. For percent operating accruals, 43 percent remain in the first two deciles in the following year. Thus, while the percent operating accruals measure picks out more extreme differences in forecasts between the naïve and sophisticated models, the accruals for these observations do not reverse any faster than for the traditional operating accruals measure.

Greater Capital Intensity? Another possibility is that, by not scaling by total assets, the percent operating accrual measure is picking up more capital-intensive firms. This is not the case, however. The lowest decile of traditional operating accruals has a ratio of depreciation/amortization to average total assets of 0.1261, while the same measure is only 0.0684 for the lowest decile of percent operating accruals.

More Extreme “Outlier” Returns? Kraft et al. (2006) show that a large part of the positive excess return in the low-accrual portfolio is due to a few firms with extreme positive returns. The 99th percentile of excess returns in the first decile of traditional operating accruals is 5.23, but is only 2.87 in the first decile of percent operating accruals. Further, the mean of the top five excess returns is almost twice as high in the first decile of traditional operating accruals than it is in the first decile of percent operating accruals. While the distribution of returns is clearly skewed in both samples, the percent operating accrual measure is less sensitive to extreme returns than is the traditional measure. Incidentally, of the 83 most extreme excess returns in the top percentile of the first decile of traditional operating accruals and percent operating accruals, only 4 observations are common between the two portfolios.

VI. CONCLUSION

Accruals are the main work product of financial accounting. It is therefore vitally important that we understand how accruals relate to equity valuation. This exploration goes beyond attempts to refine an investment strategy and produce ever-bigger excess returns. By documenting another means of using accrual information to identify misvalued stocks, we learn more about what the market does and does not understand.

We offer a subtle innovation. By redefining accruals to be relative to income rather than to total assets—that is, percent accruals—we offer a new measure of accruals that selects radically different firms and produces excess returns in subsamples of the population for which the traditional accruals measure performs poorly. Our results are consistent with the earnings fixation hypothesis insofar as extreme values of percent accruals correspond with cases in which the

difference between a naïve and sophisticated forecasting model is the most pronounced. As the literature continues to investigate the underlying explanation for the accruals anomaly, our results provide additional regularities that the theory must accommodate.

REFERENCES

- Barber, B., J. Lyon, and C. Tsai. 1999. Improved methods for tests of long-run abnormal stock returns. *The Journal of Finance* 54 (February): 165–201.
- Barth, M. E., and S. Kallapur. 1996. The effects of cross-sectional scale differences on regression results in empirical accounting research. *Contemporary Accounting Research* 13 (2): 527–567.
- Bartov, E., S. Radhakrishnan, and I. Krinsky. 2000. Investor sophistication and patterns in stock returns after earnings announcements. *The Accounting Review* 75 (1): 43–63.
- Beaver, W., M. McNichols, and R. Price. 2007. Delisting returns and their effect on accounting-based market anomalies. *Journal of Accounting and Economics* 43 (2–3): 341–368.
- Beneish, M., and M. Vargus. 2002. Insider trading, earnings quality and accrual mispricing. *The Accounting Review* 77 (4): 755–791.
- Bradshaw, M., S. Richardson, and R. Sloan. 2001. Do analysts and auditors use information in accruals? *Journal of Accounting Research* 39 (1): 45–74.
- Burgstahler, D., and I. Dichev. 1997. Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics* 24 (1): 99–126.
- Cheng, A., and W. Thomas. 2005. The effects of alternative models and estimation procedures on the abnormal accrual anomaly. Working paper, University of Houston.
- Collins, D., and P. Hribar. 2002. Errors in estimating accruals: Implications for empirical research. *Journal of Accounting Research* 40 (March): 105–134.
- D'Avolio, G., E. Gildor, and A. Shleifer. 2002. *Technology, Information Production, and Market Efficiency: Economic Policy for the Information Economy*. A Symposium Sponsored by the Federal Reserve Bank of Kansas City, KS.
- Dechow, P., and W. Ge. 2006. The persistence of earnings and cash flows and the role of special items: Implications for the accrual anomaly. *Review of Accounting Studies* 11 (2–3): 253–296.
- Dopuch, N., C. Seethamraju, and W. Xu. 2009. The pricing of accruals for profit and loss firms. *Review of Quantitative Finance and Accounting* 34, (4): 505–516.
- Desai, H., S. Rajgopal and M. Venkatachalam. 2004. Value-glamour and accruals mispricing: One anomaly or two? *The Accounting Review* 79 (2): 355–85.
- Fairfield, P., S. Whisenant, and T. Yohn. 2003. Accrued earnings and growth: Implications for future profitability and market mispricing. *The Accounting Review* 78 (1): 353–371.
- Houge, T., and T. Loughran. 2000. Cash flow is king? Cognitive errors by investors. *The Journal of Psychology and Financial Markets* 1 (March): 161–175.
- Kothari, S., E. Loutskina, and V. Nikolaev. 2006. Agency theory of overvalued equity as an explanation for the accrual anomaly. Working paper, Massachusetts Institute of Technology.
- Kraft, A., A. Leone, and C. Wasley. 2006. An analysis of the theories and explanations offered for the mispricing of accruals and accrual components. *Journal of Accounting Research* 44 (2): 297–339.
- Lakonishok, J., A. Shleifer, and R. Vishny. 1994. Contrarian investment, extrapolation, and risk. *The Journal of Finance* 49 (5): 1541–1578.
- Lesmond, D., and W. Wang. 2006. The long and short of accrual based trading strategies. Working paper, Tulane University.
- Mashruwala, C., S. Rajgopal, and T. Shevlin. 2006. Why is the accrual anomaly not arbitrated away? The role of idiosyncratic risk and transaction costs. *Journal of Accounting and Economics* 42 (1–2): 3–33.
- Richardson, S., R. Sloan, M. Soliman, and I. Tuna. 2005. Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics* 39 (3): 437–485.
- Shumway, T. 1997. The delisting bias in CRSP data. *The Journal of Finance* 52 (1): 327–340.
- , and V. A. Warther. 1999. The delisting bias in CRSP's Nasdaq data and its implications for the size effect. *The Journal of Finance* 54 (6): 2361–2379.

- Sloan, R. 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71 (July): 289–316.
- Zach, T. 2004. Evaluating the “accrual-fixation” hypothesis as an explanation for the accrual anomaly. Working paper, Washington University.

Do Investors Understand *Really Dirty* Surplus?

Wayne R. Landsman

The University of North Carolina at Chapel Hill

Bruce L. Miller

University of California, Los Angeles

Ken Peasnell

Lancaster University

Shu Yeh

National Taiwan University

ABSTRACT: This study addresses whether firms' share prices correctly reflect two accounting measures: dirty surplus and really dirty surplus. Dirty surplus is readily observable from the financial statements, but really dirty surplus, which arises from recognizing equity transactions such as employee stock option exercises at other than fair market value, is not. Findings show that dirty surplus and really dirty surplus are irrelevant for forecasting abnormal comprehensive income. However, findings also indicate that investors appear to undervalue really dirty surplus. Hedge returns are insignificant when portfolios are formed based on dirty surplus, but are significantly positive based on really dirty surplus. Really dirty surplus positive hedge returns are robust to a variety of sensitivity tests. Taken together, the findings are consistent with either investors over-valuing firms that have large negative really dirty surplus or really dirty surplus being correlated with an unmodeled risk factor.

Keywords: *dirty surplus accounting; equity valuation; hedge returns.*

Data Availability: *Data are available from sources identified in the paper.*

We thank workshop participants at the 2009 European Accounting Association Congress, 2008 UNC-Duke Fall Camp, 2008 Yale Fall Camp, Chinese University of Hong Kong, National Taiwan University, National University of Singapore, Norwegian School of Economics, University of California, Los Angeles, University of Graz, and Vienna University of Economics and Business Administration, as well as Frøystein Gjesdal and two anonymous reviewers for helpful comments. We acknowledge funding from the Center for Finance and Accounting Research at The University of North Carolina at Chapel Hill and National Science Council, Republic of China (NSC 96-2416-H-002-038; NSC97-2410-H-002-042).

Editor's note: Accepted by Sanjay Kallapur.

Submitted: July 2008

Accepted: July 2010

Published Online: January 2011

I. INTRODUCTION

A substantial and growing literature considers whether investors properly assess the characteristics of earnings and its components when setting stock prices. The question we address is whether firms' share prices correctly reflect two accounting measures that have received relatively little attention to date. The first of these is commonly referred to as "dirty surplus," which is a component of comprehensive income that is excluded from reported earnings, and therefore violates clean surplus accounting. We label the second accounting measure we consider "really dirty surplus," which arises when a firm issues or reacquires its own shares in a transaction that does not record the shares at fair market value. Examples of this kind of transaction are shares issued in a stock option exercise and a conversion of a bond into common stock. Prior to the implementation of FASB Statement No. 141, the pooling-of-interests method of accounting for business combinations could also result in substantial really dirty surplus. If investors fully understand the predictive value of these accounting amounts, then it should not be possible to develop a profitable trading strategy based on the magnitudes of these items.

Unlike dirty surplus, which is readily observable from the financial statements, really dirty surplus is unobservable. That is, even the most sophisticated investor cannot estimate readily the valuation impact of equity transactions that give rise to really dirty surplus because equity transactions are recognized in the financial statements using an accounting-based rather than a market-based measure of the value of equity. The estimation task investors face is exacerbated by the fact that a firm can engage in numerous such transactions throughout the year. As a result, really dirty surplus transactions are less likely to be correctly priced than are dirty surplus transactions.

Using financial statement and stock price data from 1976–2006, we first assess whether investors properly value each of these two components of earnings by estimating a residual income forecasting equation and an attendant valuation equation that includes both of these components. If current residual income is sufficient for forecasting the next period's residual income and current residual income and equity book value are sufficient for valuing current equity, then the forecasting and valuation coefficients on the income components of interest will be linked in a predictable manner. Finding a mismatch between the components' forecasting and valuation equation coefficients would be consistent with investors' mispricing of the components.

We also conduct hedge portfolio returns tests. We adopt a buy-and-hold strategy to go long in firms with relatively large dirty or really dirty surplus and to go short in firms with relatively small dirty or really dirty surplus. We conjecture that small firms' prices are less likely to reflect fully all publicly available information because investors incur proportionately greater transaction costs; as a result, they are less closely followed and less likely to be subject to detailed accounting analysis. We therefore conduct both sets of tests separately for small, medium, and large firms to assess whether pricing effects are related to firm size.

We find that both dirty surplus and really dirty surplus are irrelevant for forecasting abnormal very comprehensive income for all three firm-size groups. Taking these results at face value, if investors correctly understand the implications of these persistence findings for valuation, then each kind of dirty surplus should be irrelevant for valuation for all firms. This prediction is borne out in the case of dirty surplus. However, the findings indicate that investors appear to undervalue really dirty surplus, which is consistent with investors being unable to assess the economic implications of really dirty surplus transactions.

Buy-and-hold hedge return results support the findings from the tests linking the forecasting and valuation equations. As expected, hedge returns are insignificantly different from zero when based upon dirty surplus, regardless of firm size and investing horizon. In contrast, the hedge returns based on really dirty surplus are significantly positive for all three firm-size groups. We also consider an alternative to our buy-and-hold procedure for computing hedge returns. Findings based on mean returns for monthly calendar-time hedge portfolios indicate that significantly posi-

tive hedge returns are concentrated within small firms. Findings from additional tests reveal that inferences relating to hedge returns are insensitive to including controls for four previously identified mispricing anomalies, and to sampling procedures designed to attempt to focus on sources of really dirty surplus.

Taken together, the findings are consistent with investors over-valuing firms that have large negative really dirty surplus. However, several cautionary notes are in order. First, although the hedge returns findings are consistent with mispricing of really dirty surplus, the possibility remains that the mismatch of the really dirty surplus forecasting and valuation coefficients is the result of model misspecification rather than mispricing. Second, as is likely the case with investors, we are unable to trace the sources of really dirty surplus to particular types of equity transactions. As a result, we cannot determine the extent to which potential mispricing arises from each type of transaction, i.e., our findings can only be interpreted as reflecting the aggregate effect of the various types of transactions. However, even if we could trace the sources of really dirty surplus, any resulting hedge returns might still be attributable to an unmodeled risk factor.

Our study adds to prior research finding evidence of investors' apparent failure to link the forecasting attributes of accounting amounts with the pricing implications (e.g., Bernard and Thomas 1989; Sloan 1996; Barth et al. 1999; Bradshaw and Sloan 2002; Burgstahler et al. 2002; Brown and Sivakumar 2003; Doyle et al. 2003; Landsman et al. 2007). Our findings support prior studies that find that investors understand the forecasting properties and valuation implications of dirty surplus (Dhaliwal et al. 1999; O'Hanlon and Pope 1999; Biddle and Choi 2006; Chambers et al. 2007). Although Landsman et al. (2006) examine valuation implications of expected future equity transactions arising from the exercise of employee stock options, their study does not address whether investors take full account of the valuation implications of past option exercises. Core et al. (2002) report findings suggesting that dilutive transactions, including those arising from employee stock option grants, are poorly dealt with in reported diluted earnings per share, leaving open the possibility that investors may have difficulty in valuing such transactions.

Section II provides the motivation for the study and explains how dirty surplus and really dirty surplus are defined. Section III presents the research design, including computation of dirty surplus and really dirty surplus, development of the forecasting and valuation equations, and description of our hedge return strategy. Section IV describes the sample and data, and Section V presents the findings. Section VI summarizes and concludes.

II. MOTIVATION

The empirical issue that is central to this research is whether firms' share prices correctly reflect two accounting measures that have received relatively little attention to date. The first of these is commonly referred to as *dirty surplus* (*DS*), which is a component of comprehensive income that is excluded from reported earnings and therefore violates clean surplus accounting (Ohlson 1995; Feltham and Ohlson 1995). Dirty surplus accounting results in the basic residual income valuation model yielding an inaccurate estimate of equity value because the sum of current book value and future net incomes does not equal the sum of future net dividends.

The second accounting measure we consider we label *really dirty surplus* (*RDS*), which arises when a firm issues or reacquires its own shares in a transaction that does not record the shares at fair market value. The primary sources of *RDS* include employee stock option exercises, conversion of preferred stock and bonds, and mergers accounted for as pooling-of-interests. Whereas equity issued under employee stock option exercises or convertible instruments can give rise to unrecorded expenses, equity issued under pooling-of-interests gives rise to an unrecorded asset.¹

¹ Equity issued under the pooling-of-interests method is not recognized at fair market value. In contrast, if purchase

RDS violates the *super-clean surplus* concept (Feltham 1996; Christensen and Feltham 2003), under which it is assumed that share issuances are recorded at fair market value. When this condition is violated, the discounted present value of future net dividends (or equivalently, the sum of equity book value and discounted present value of future abnormal earnings) will not equal the market value of equity relating to current shares outstanding, but rather will equal the market value of equity relating to current shares outstanding plus the market value of other equity claimants. Because the equity transactions that give rise to *RDS* generally are recorded in the financial statements at less than market value, *RDS* is generally negative.²

DS is readily observable from the financial statements. When one takes a *clean surplus accounting* perspective, as comprehensive income does, *dirty surplus* becomes a component of earnings. Dirty surplus is conventionally defined as the sum of recognized revenue or expense items that bypass the income statement. Unlike *DS*, *RDS* is not reported in the financial statements. A *super-clean surplus accounting* perspective requires that both *dirty surplus* and *really dirty surplus* become components of earnings so that the discounted stream of future residual incomes and current equity book sums to the market value of equity of current shareholders. If investors fully understand the implications of *DS* and *RDS* for valuation, then it should not be possible to develop a profitable trading strategy based on the magnitudes of these items.

To see these points more clearly, consider first the following version of clean surplus accounting:

$$BVE_t = BVE_{t-1} + X_t + DS_t - Div_t + P_t^A(N_t - N_{t-1}), \quad (1)$$

where BVE_t is defined as ending equity book value, X_t represents net income, DS_t is dirty surplus, Div_t is dividends, N_t is the number of shares outstanding at the end of period t , and P_t^A is the price per share used to record the issuance or reacquisition of equity shares in the accounting system. Note that if DS_t is zero, then the accounting is said to satisfy clean surplus accounting.

Let P_t^M be the market price per share at the date of issuance or reacquisition of equity shares. We define *really dirty surplus*, RDS_t , by:

$$RDS_t \equiv (N_t - N_{t-1})(P_t^A - P_t^M). \quad (2)$$

By combining Equation (1) and Equation (2), we arrive at:

$$DS_t + RDS_t = BVE_t - BVE_{t-1} - X_t + Div_t - P_t^M(N_t - N_{t-1}). \quad (3)$$

DS_t , BVE_t , BVE_{t-1} , X_t , and Div_t are readily observable in the financial statements. The final term on the right-hand side of Equation (3), $P_t^M(N_t - N_{t-1})$, is not reported in the financial statements and therefore needs to be estimated.

Note that if both DS_t and RDS_t are zero in Equation (3), then the accounting is said to satisfy *super-clean surplus accounting*. The next section allows for both nonzero DS_t and RDS_t and *super-clean surplus accounting* by setting $VCNI_t = X_t + DS_t + RDS_t$, where *VCNI* is “very comprehensive” net income. Our definition of *RDS*, and hence *VCNI*, attributes all of the violation of *super-clean surplus accounting* to the period during which the equity transaction is recorded at a price other than fair market value. Christensen and Feltham (2003) show that when *super-clean surplus accounting* holds in periods subsequent to time t , application of the residual income

accounting was applied instead of the pooling-of-interests method, *RDS* would not arise; instead, the amount of *RDS* attributed to pooling-of-interests would be recognized in the financial statements as goodwill.

² It is possible that *RDS* could be positive. For example, consider a bond with a book value of \$175 and fair value of \$125 that is converted into equity whose fair value is \$150. In this case, *RDS* would be a positive amount equal to \$25. Likewise, unrecorded goodwill associated with a merger accounted for under pooling-of-interests could give rise to negative unrecorded goodwill and, hence, positive *RDS*.

valuation model will yield an estimate of equity value that equals the market value of equity of existing shares. Whether super-clean surplus accounting holds up to and including period t simply affects the opening balance of equity book value at time t . However, if P^M and P^A differ for transactions in periods subsequent to time t , then super-clean surplus accounting will be violated and, hence, the residual income valuation model will not yield an estimate of equity value that equals the market value of existing shares.³

As stated above, the empirical issue that is central to this research is whether firms' price per share correctly reflects DS and RDS . If it does, then one should not be able to develop a trading strategy based on DS or RDS that generates future abnormal returns. There are several reasons why we expect that RDS is the better earnings component on which to base a trading strategy. First, as noted above, unlike DS , RDS is not reported in the financial statements. Second, RDS appears to be inherently complex. For example, for most earnings components, any "overstatement" or "understatement" reverses in future periods; this does not hold for RDS . Third, research on DS (Dhaliwal et al. 1999; O'Hanlon and Pope 1999; Biddle and Choi 2006; Chambers et al. 2007) is not especially encouraging about the possibility that it can be used to construct a profitable trading strategy.

Nonetheless, there are at least two compelling reasons for conducting our tests for DS as well as RDS . First, our study is the first to examine whether investors properly price DS based on the forecasting and valuation equations in the Ohlson (1999) model as well as on hedge return tests. Second, because we do not necessarily expect to find evidence of mispricing relating to DS , finding this is the case mitigates concerns that finding evidence of mispricing relating to RDS is attributable to misspecification of our empirical procedures.

III. RESEARCH DESIGN

Computation of DS and RDS

Following Dhaliwal et al. (1999) and Chambers et al. (2007), we compute DS as the sum of (1) the change in the balance of unrealized gains or losses on marketable securities (change in Compustat #238), (2) the change in the cumulative foreign exchange adjustment (change in Compustat #230), and (3) 0.65 times the change in additional pension liability in excess of unrecognized prior service costs (change in Min [(Compustat #297 - #298), 0]).

Based on Equation (3), we compute RDS as the change in the book value of common equity (Compustat #60 + #227 - #242), less DS , less net income (Compustat #172 - #19), plus dividends (Compustat #21), less share price at middle of fiscal year times change in common shares outstanding (Compustat #25, adjusted for stock dividends and splits).⁴ Note that because we (and investors) cannot readily compute using the individual underlying equity transactions, RDS likely measures the true underlying construct with error.⁵ Share prices are from the CRSP database.

³ Landsman et al. (2006) show that, in the case of employee stock options (i.e., contingent equity), when only clean surplus holds, the estimate of equity value equals the sum of the market value of existing shares and employee stock options. The study's model considers the case in which employee stock options are granted at time t or earlier. The residual income valuation model does not yield an estimate of the value of existing shares because the options are not yet exercised, and when they are exercised in the future, the new shares will be recognized at P^A rather than P^M .

⁴ To the extent that our definition of DS does not include all dirty surplus items (e.g., the cumulative effects on equity of retrospective accounting changes), DS will be measured with error. Because RDS is net of DS , such items will appear as part of our measure of RDS . Also, our measure of RDS includes treasury stock transactions taking place at prices that differ from the market price at the middle of the fiscal year.

⁵ In particular, the use of mid-year prices in the construction of RDS is arbitrary. We test the sensitivity of our findings to measuring RDS at alternative dates using both end-of-year and average of beginning- and end-of-year prices. Untabulated findings based on these alternative measures reveal that none of the inferences are affected.

Forecasting and Valuation Equations

To examine how the dirty surplus and really dirty surplus components of income relate to equity value, we adopt the abnormal earnings forecasting and equity valuation equations from Barth et al. (1999), which are based on the linear information system developed in Ohlson (1999):

$$VCNI_{it+1}^a = \omega_0 + \omega_1 VCNI_{it}^a + \omega_2 DS_{it} + \omega_3 RDS_{it} + \omega_4 BVE_{it} + \varepsilon_{it+1}. \quad (4)$$

$$MVE_{it} = \alpha_0 + \alpha_1 VCNI_{it}^a + \alpha_2 DS_{it} + \alpha_3 RDS_{it} + \alpha_4 BVE_{it} + u_{it}. \quad (5)$$

Equation (4) is the abnormal earnings forecasting equation, where abnormal very comprehensive earnings, $VCNI_t^a$, is defined as very comprehensive earnings, $VCNI_t$, less a normal return on beginning equity book value, BVE_{t-1} , i.e., $VCNI_t - rBVE_{t-1}$. Very comprehensive income is net income, NI_t , plus both dirty surplus and really dirty surplus. Following Ohlson (1999) and Barth et al. (1999), $VCNI_t$ is partitioned into NI_t , DS_t , and RDS_t . The linear information system represented by Equation (4) and Equation (5) implicitly assumes that current earnings amounts are predictive of all future earnings. To the extent that this assumption is violated, the algebraic links between forecasting coefficients in Equation (4) and the valuation coefficients in Equation (5) described below do not necessarily hold. Of particular significance to this study is whether current realizations of DS and RDS are predictive of future $VCNI^a$ that includes future realizations of these variables.

In Equation (4), ω_1 reflects the persistence of abnormal earnings. Prior research (e.g., Dechow et al. 1999; Barth et al. 1999, 2005) leads us to predict that ω_1 is positive.⁶ The coefficients on the DS and RDS earnings components, ω_2 and ω_3 , reflect the incremental effects on the forecast of abnormal earnings of knowing these components. If all earnings components have the same ability to forecast $VCNI_{t+1}^a$, then ω_2 and ω_3 will both equal zero; thus, knowing each component of earnings does not aid in forecasting abnormal earnings. As a result, we test the null hypotheses that $\omega_2 = 0$ and $\omega_3 = 0$ against the alternative that $\omega_2 \neq 0$ and $\omega_3 \neq 0$.

Following Ohlson (1999, 150), we define DS (RDS) as being “forecasting-irrelevant” if the quadruple $\{NI_t, RDS_t, BVE_t, BVE_{t-1}\}$ ($\{NI_t, DS_t, BVE_t, BVE_{t-1}\}$) contains the same information as the quintuple $\{NI_t, DS_t, RDS_t, BVE_t, BVE_{t-1}\}$ for purposes of forecasting $VCNI_{t+1}^a$. Because DS and RDS are components of $VCNI_t^a$, the total coefficients on DS_t and RDS_t are $\omega_1 + \omega_2$ and $\omega_1 + \omega_3$. ω_4 is not included in the total coefficient on either DS_t or RDS_t because BVE_t is unchanged across the different definitions of clean surplus and is therefore invariant to the definition of clean surplus. Thus, if $\omega_1 + \omega_2 = 0$ ($\omega_1 + \omega_3 = 0$), DS (RDS) is irrelevant for forecasting abnormal earnings. Conversely, if $\omega_1 + \omega_2 \neq 0$ ($\omega_1 + \omega_3 \neq 0$), then DS (RDS) is said to have abnormal earnings “forecasting relevance.” To examine whether dirty surplus and really dirty surplus components of comprehensive income are forecasting-irrelevant, we test the null hypotheses that $\omega_1 + \omega_2 = 0$ and $\omega_1 + \omega_3 = 0$ against the alternatives that $\omega_1 + \omega_2 \neq 0$ and $\omega_1 + \omega_3 \neq 0$. Note that ω_1 reflects the forecasting relevance of the $VCNI_t^a - DS_t - RDS_t = NI_t - rBVE_{t-1}$ component of $VCNI_t^a$.

Equation (5) is the valuation equation based on the information dynamics in Equation (4). α_2 and α_3 , the valuation multiples on DS and RDS , can be interpreted in a symmetrical fashion. This follows from the fact that although DS_t is by definition included in BVE_t , it follows from Equation (1) and Equation (3) that RDS_t that arises from dilutive transactions is normally included in BVE_t .

⁶ Ohlson (1995, 1999) permits the forecasting and valuation equations to include “other information.” Fairfield et al. (2003) show that accruals and asset growth have incremental ability to predict future return on assets. Accordingly, viewing accruals and asset growth as “other information,” below we report findings from alternative specifications of Equation (4) and Equation (5) that include proxies for these variables as additional explanatory variables.

as well.⁷ Analogous to the interpretation of ω_2 (ω_3) in Equation (4), a_2 (a_3) reflects the incremental effect on valuation from knowing DS (RDS). If all earnings components are equally persistent, then they should have the same relation with equity value. If this is the case, then a_2 and a_3 will equal zero, and knowing each component of earnings will not aid in explaining equity value. Thus, we test the null hypothesis that $a_2 = 0$ ($a_3 = 0$) against the alternative that $a_2 \neq 0$ ($a_3 \neq 0$). We define DS (RDS) as being “valuation-irrelevant” if the quadruple $\{NI_t, RDS_t, BVE_t, BVE_{t-1}\}$ ($\{NI_t, DS_t, BVE_t, BVE_{t-1}\}$) contains the same information as the quintuple $\{NI_t, DS_t, RDS_t, BVE_t, BVE_{t-1}\}$ for purposes of valuation. Also analogous to Equation (4), the total valuation coefficient on DS (RDS) equals $a_1 + a_2$ ($a_1 + a_3$). Thus, if $\alpha_1 + \alpha_2 = 0$ ($\alpha_1 + \alpha_3 = 0$), DS (RDS) is irrelevant for valuation.⁸ Conversely, if $a_1 + a_2 \neq 0$ ($a_1 + a_3 \neq 0$), then DS (RDS) is “valuation-relevant.” Analogous to the interpretation of ω_1 in Equation (4), a_1 reflects the value relevance of the $VCNI_t^a - DS_t - RDS_t = NI_t - rBVE_{t-1}$ component of $VCNI_t^a$.

Barth et al. (1999) derive a formula linking the coefficients in Equation (4) and the two suppressed equations with the coefficients in Equation (5). For our purposes, we are not interested in exact coefficient magnitudes based on imposing a full set of linear information dynamics. Instead, we are interested in the weaker prediction that the sign of $a_1 + a_2$ ($a_1 + a_3$) will be based on the sign of $\omega_1 + \omega_2$ ($\omega_1 + \omega_3$).

If prices are determined rationally, then if DS or RDS is irrelevant for forecasting the next period amount, each should be valuation-irrelevant as well if the linear dynamics in Equation (4) and Equation (5) hold. Also, the sign of $a_1 + a_2$ ($a_1 + a_3$) will be the same as the sign of $\omega_1 + \omega_2$ ($\omega_1 + \omega_3$). If we find apparent evidence of mispricing based on the empirical coefficients from estimating Equation (4) and Equation (5), then a buy-and-hold strategy of going long in relatively underpriced stocks and short in relatively overpriced stocks should yield excess returns.

Any mismatch between the forecasting and valuation results for DS or RDS need not necessarily be attributable to its mispricing by investors. It might be the case, for example, that although RDS_t cannot be used to forecast $VCNI_{t+1}^a$, it could be used to forecast $VCNI_{t+k}^a$ ($k = 2, 3, \dots$). If this were the case, then RDS would not be valuation-irrelevant, and the mismatch between forecasting and valuation coefficients would be attributable to variables omitted from both equations. The hedge returns tests provide a means of examining this issue.

If transaction cost considerations imply that small firms are more difficult to price, then we would expect the hedge portfolio returns to be greater than in the case of larger firms. Therefore, we estimate and test predictions relating to Equation (4) and Equation (5) separately for small, medium, and large firms based on equity market value and conduct hedge return tests also separately for small, medium, and large firms.

⁷ This can be illustrated by the following simple bond conversion example. Consider a firm that has a convertible bond outstanding on its books at \$100 that is converted into shares worth \$150 at time t . Under current GAAP, the share issuance will be recorded at \$100. If we assume for simplicity that $X_t = DS_t = 0$ and that $BVE_{t-1} = \$1,000$, it follows that $BVE_t = 1,000 + 100 = \$1,100$. If this transaction were to be accounted for on a super-clean surplus basis, the share issuance would be recorded at \$150, with the resultant cost of conversion appearing as $RDS_t = -50$. We can deduce from Equation (3) that under super-clean surplus accounting $BVE_t = 1,000 - 50 + 150 = \$1,100$ as well. Although the calculations are more complex in the case of employee stock options, the same conclusion applies. Note that in the case of mergers accounted for under pooling-of-interests, the inclusion of RDS_t in BVE_t would leave BVE_t unchanged only if the asset (goodwill) associated with an acquisition was immediately expensed.

⁸ Note that under the Ohlson (1999) framework, value irrelevance (e.g., which occurs for DS (RDS) when $\alpha_1 + \alpha_2 = 0$ ($\alpha_1 + \alpha_3 = 0$)) of an earnings component implies that it has no impact on goodwill, which is the difference between equity market value and book value. Ohlson (1999, 152) further states that “an incremental dollar of transitory earnings adds a dollar to market value. This claim is easy to validate as long as one keeps in mind that a dollar of transitory earnings also adds a dollar to book value.”

Hedge Portfolio Strategy and Procedure

Hedge Strategy Overview

We determine the hedge portfolio strategy in the following manner. First, for each sample year, we rank firms according to either *DS* or *RDS* as a fraction of end-of-year equity book value, *BVE*.⁹ We then form ten portfolios whereby the first (tenth) portfolio contains those observations with the smallest (largest) fraction of *DS* or *RDS*. Second, within each of the ten *DS* or *RDS* portfolios we rank firms according to equity market capitalization and assign each firm to one of three equal-sized groups of firms comprising the small, medium, and large firms. This procedure results in there being ten portfolios within each of the three firm-size groups. It also ensures that the magnitude of *DS* or *RDS* does not vary systematically across the three firm-size groups, and thereby helps to mitigate the confounding effect of firm size when conducting our hedge portfolio tests.¹⁰ Third, we then combine observations from all sample years, retaining the firm size designation and *DS* and *RDS* portfolio rankings. This results in there being three firm-size groups, within each of which there are ten *DS* or *RDS* portfolios.¹¹ Fourth, within each of the three firm-size groups, for each firm in the ten *DS* or *RDS* portfolios, we compute the risk-adjusted return over all sample years. Fifth, we compute the hedge return by deducting the equally weighted mean risk-adjusted return on the portfolio(s) comprising firms we expect to be most over-valued from the return on portfolio(s) comprising firms we expect to be either undervalued or least over-valued.¹²

We predict that over-valuation is most likely to occur for firms whose income is overstated relative to very comprehensive net income, and where the market fails to understand the economic implications of such overstatement. As noted in Section II, we expect these conditions to be more descriptive for *RDS* than *DS*. Recall that *RDS* is generally non-positive because the accounting procedures that give rise to *RDS* arise from equity transactions that generally are recorded at less than market value. Our hedge strategy is therefore long in firms with least negative *RDS* and short in firms with most negative *RDS*. We employ a similar strategy for *DS*, i.e., go long in firms with most positive *DS* and short in firms with most negative *DS*. As noted above, we do not expect this *DS*-based hedge strategy to yield significant positive (or negative) excess returns.

Following Bernard and Thomas (1990), we compute the hedge return for each of the three firm-size groups by going long (short) in the firms in the top three (bottom three) *DS* or *RDS* portfolios. Combining observations in the top three and bottom three *DS* or *RDS* portfolios confers the benefit of mitigating the potential effects of measurement error in the extreme *DS* or *RDS* portfolios. We employ the hedge portfolio tests to complement the tests based on the forecasting and valuation equations. In particular, if the forecasting and valuation equations yield evidence of mispricing, notably undervaluation of *DS* or *RDS*, then the hedge portfolio tests should yield evidence that excess returns can be earned by exploiting such undervaluation. Conversely, if the forecasting and valuation equations yield no evidence of mispricing, then the hedge portfolio tests

⁹ Untabulated findings based on *DS* and *RDS* deflated by total assets result in no changes in inferences.

¹⁰ By design, this procedure is a double-conditional sort of first *RDS* (or *DS*) then size. As a consequence, this procedure can fail to adequately control for size differences between long and short *RDS* portfolios. To assess the sensitivity of hedge returns, we reversed the sorting procedure and recomputed hedge returns sorting firm-years using a double-conditional sort of first size then *RDS*. Hedge return findings based on this alternative procedure result in inferences that are substantially the same as those based on the tabulated hedge return findings.

¹¹ Because firm size is increasing during our sample period, some large firms in early sample years would be considered small firms in later sample years. However, because firm-size groupings are determined annually, our procedure mitigates year effects on our hedge portfolio test inferences.

¹² Untabulated findings based on hedge returns computed with value-weighted portfolio risk-adjusted returns result in no change in inferences. Additional untabulated findings based on cumulative abnormal returns also result in no change in inferences.

should yield evidence that excess returns cannot be earned following our hedge strategy.

Hedge Strategy Implementation Details

To estimate risk-adjusted return, we need a measure of expected stock return. Following Ang and Liu (2004), Ibbotson and Associates (2005), Massa et al. (2005), and Barth et al. (2008), we use the Fama and French (1993) three-factor model, supplemented with the momentum factor (Jegadeesh and Titman 1993; Carhart 1997), with time-varying factor loadings, risk-free rates, and risk premia. We calculate each firm's expected equity return for month $t+1$ as of month t , $ER_{i,t+1}$, conditional on the expected factor returns in month $t+1$, based on Equation (6):

$$ER_{i,t+1} = R_{f,t+1} + \beta_{RMRF,i,t+1}(R_{M,t+1} - R_{f,t+1}) + \beta_{SMB,i,t+1}SMB_{t+1} + \beta_{HML,i,t+1}HML_{t+1} + \beta_{MOM,i,t+1}MOM_{t+1}, \quad (6)$$

where $\beta_{RMRF,i,t+1}$, $\beta_{SMB,i,t+1}$, $\beta_{HML,i,t+1}$, and $\beta_{MOM,i,t+1}$ are firm-specific coefficients estimated from Equation (7) below. $R_{M,t+1} - R_{f,t+1}$, SMB_{t+1} , HML_{t+1} , and MOM_{t+1} are the expected monthly Fama-French and momentum factor returns for month $t+1$. We estimate the expected monthly factor returns for month t by calculating each factor's average monthly return over the 60 months prior to month t . The difference, $R_{M,t} - R_{f,t}$, is the monthly return of the market portfolio in excess of the risk-free rate, HML_t and SMB_t are the monthly returns to the book-to-market and size factor-mimicking portfolios, respectively, as described in Fama and French (1993), and MOM_t is the monthly return to the momentum factor-mimicking portfolio. The risk-adjusted return for firm i in month $t+1$ is the difference between the firm's realized return in month $t+1$, $R_{i,t+1}$, and its expected return, $ER_{i,t+1}$. We then use these monthly risk-adjusted returns to compute annual returns. In the hedge return tests, we cumulate return three months after fiscal year-end to ensure that the financial statements are available to the public.

For each firm, we estimate the betas associated with the firm's return to each of the Fama-French and momentum factors by estimating the following monthly time-series regression:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{RMRF,i}(R_{M,t} - R_{f,t}) + \beta_{SMB,i}SMB_t + \beta_{HML,i}HML_t + \beta_{MOM,i}MOM_t + \varepsilon_{i,t}, \quad (7)$$

where $R_{i,t} - R_{f,t}$ is the firm's monthly return in excess of the risk-free rate. We estimate Equation (7) using the most recent 60-month returns prior to the month t . This results in estimated coefficients, $\hat{\beta}_{RMRF,i,t}$, $\hat{\beta}_{SMB,i,t}$, $\hat{\beta}_{HML,i,t}$, and $\hat{\beta}_{MOM,i,t}$, which are updated monthly. We define our forecast of each factor beta for month $t+1$ using the fitted value for that factor for month t , e.g., $\beta_{RMRF,i,t+1} = \hat{\beta}_{RMRF,i,t}$.

Following Doyle et al. (2003), we compute hedge returns over one-, two-, and three-year horizons. We conjecture that if hedge returns continue to increase over longer horizons, then such evidence would be indicative of unmodeled risk differences. Therefore, we expect hedge returns to flatten over the three-year horizon. To avoid imposing the assumption of normality of the distribution of excess returns, we report an additional test for significance of the hedge returns using a t-test based on a bootstrapping procedure. Specifically, we select firm observations that we randomly assign to the ten portfolios. We then calculate the hedge return. We repeat this procedure 1,000 times, thereby generating an empirical distribution that we use to report empirical p-values in addition to conventional t-statistics and their implied p-values.

IV. SAMPLE AND DESCRIPTIVE STATISTICS

We obtain most of the data for estimation of Equation (4) and Equation (5) for 1976–2006 from the Compustat Primary, Secondary, and Tertiary, Full Coverage, and Research Annual Industrial Files. DS_t and RDS_t are calculated using Compustat and CRSP data as described in the "Computation of DS and RDS " section. We compute $VCNI_t^a$ as $VCNI_t - rBVE_{t-1}$, where $VCNI_t$

includes both DS_t and RDS_t . Following Barth et al. (1999, 2005), Dechow et al. (1999), Bell et al. (2002), and Landsman et al. (2007), we set r , the cost-of-equity capital, equal to 12 percent, and we require sample firms to have positive equity book value.¹³ We also require that sample firms have total assets in excess of \$10 million to avoid the undue influence of small firms. To mitigate the effects of outliers, for each variable within each of the three size categories, we treat as missing observations that are in the extreme top and bottom one percentile. For each sample year, firms are ranked according to end-of-year market value of common equity and assigned into one of three equal-sized groups of firms comprising the small, medium, and large firms. We estimate Equation (4) and Equation (5) using unscaled data (Barth and Kallapur 1996). We assess significance of regression coefficients using two-way clustered standard errors, with firm and year clusters (Petersen 2009; Gow et al. 2010).¹⁴ The final sample for estimation of Equation (4) and Equation (5) comprises 37,097 firm-year observations.

We obtain stock return, R , from CRSP and R_f , the one-month Treasury rate, and the Fama-French and momentum factor returns from the Fama-French database (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). To obtain excess returns per Equation (6), we estimate factor loadings from Equation (7) using monthly return data beginning in 1972.¹⁵ There are 30,383 potential DS firm-year excess return observations. However, because there are 17,579 observations with zero DS , we limit our DS hedge return analysis to the 12,804 nonzero observations. There are 28,346 RDS firm-year excess return observations.

Table 1, Panels A–D present distributional statistics and Panel E presents Pearson and Spearman correlations. Panels A–D reveal that, on average, the market value of equity exceeds the book value of equity for all size firms and mean abnormal earnings, $VCNI^a$, is positive for large firms but negative for medium and small firms. Table 1, Panel E reveals that the explanatory variables in Equation (4) and Equation (5) are correlated with each other, but not so much as to raise collinearity concerns. Although the distributional statistics reported in Panels A–D reveal the variables are skewed, none of the key inferences are affected when the equations are estimated on a per-share basis.

Because typically $P_t^M > P_t^A$ and $N_t > N_{t-1}$, we expect RDS to be negative. Table 1, Panels A–D reveal that RDS turns positive before the 75th percentile. Untabulated statistics reveal that RDS turns positive between the 60th and 70th percentiles for two-thirds of the sample years, and beyond the 70th percentile for the remaining third. Because it is unlikely that equity transactions will give rise to positive RDS , this means that at least some of our observations are measured with error. Assuming this error is unsystematic, the implication of this is a reduction in power of our tests, particularly those relating to the hedge returns.

Untabulated statistics reveal that mean RDS is economically largest (i.e., most negative) for Pharmaceuticals, Services, Food, and Computers, and that mean RDS is economically largest in 1997, 1998, and 2002. To mitigate the impact of particular industries or years overly influencing our results, below we report supplementary findings from hedge return tests that exclude those industries and years with the largest mean RDS values.

¹³ None of our inferences are affected by assuming alternative values for r , including firm-specific values based on our multi-factor model.

¹⁴ We also compute significance levels using bootstrapping. Untabulated findings result in no changes in inferences from those based on reported findings for RDS . For DS , the forecasting and valuation coefficients are still consistent, but with both significantly positive forecasting and valuation coefficients for small and medium firms.

¹⁵ Although excess returns can be computed through 2006, our sample stops in 2003. This is because we compute hedge returns for one-, two-, and three-year horizons, and to facilitate comparison over returns time, we use a common sample for the full three-year horizon.

TABLE 1

Descriptive Statistics for Equity Market Value, Equity Book Value, Abnormal Earnings, Dirty Surplus, and Really Dirty Surplus for a Sample of 37,097 Firm-Year Observations 1976–2006

Panel A: Distributional Statistics (in \$ million)					
Variable	Mean	25th %	Median	75th %	Std. Dev.
MVE	2,216.57	89.98	388.66	1,537.13	6,282.58
BVE	894.42	65.67	213.53	745.84	2,151.25
VCNI	102.05	2.01	15.50	73.56	377.09
VCNI ^a	4.43	-13.30	-0.27	12.96	262.01
DS	0.68	0.00	0.00	0.00	49.43
RDS	-18.90	-6.27	-0.42	0.02	150.51
Panel B: Distributional Statistics for Small Firms (in \$ million) ^a					
Variable	Mean	25th %	Median	75th %	Std. Dev.
MVE	85.88	24.22	50.64	99.54	102.40
BVE	64.65	20.96	40.99	81.83	66.82
VCNI	2.37	-0.18	2.46	6.51	11.83
VCNI ^a	-5.06	-6.79	-1.48	0.78	13.17
DS	0.04	0.00	0.00	0.00	1.23
RDS	-0.73	-0.52	-0.03	0.01	3.76
Panel C: Distributional Statistics for Medium Firms (in \$ million) ^a					
Variable	Mean	25th %	Median	75th %	Std. Dev.
MVE	531.34	189.03	369.80	700.67	487.35
BVE	283.47	112.81	200.98	366.31	249.43
VCNI	22.95	7.71	18.52	37.65	43.79
VCNI ^a	-7.95	-14.94	0.44	9.11	44.01
DS	0.27	0.00	0.00	0.00	6.62
RDS	-4.61	-4.94	-0.67	0.02	18.51

(continued on next page)

Panel D: Distributional Statistics for Large Firms (in \$ million)^a

Variable	Mean	25th %	Median	75th %	Std. Dev.
MVE	5,914.80	1,240.97	2,685.64	6,020.24	9,741.45
BVE	2,290.81	602.20	1,167.00	2,477.00	3,248.53
VCNI	275.33	46.54	123.25	318.75	608.33
VCN ^a	25.62	-52.27	18.65	97.62	446.07
DS	1.70	-2.05	0.00	1.34	84.46
RDS	-50.37	-36.63	-5.32	0.15	254.37

Panel E: Correlations, with Pearson (Spearman) Correlations above (below) the Diagonal

Variable	MVE	BVE	VCNI	VCN ^a	DS	RDS
MVE	1.000	0.802	0.733	0.343	0.044	-0.251
BVE	0.942	1.000	0.761	0.206	0.059	-0.202
VCNI	0.675	0.672	1.000	0.784	0.108	0.167
VCN ^a	0.186	0.105	0.628	1.000	0.118	0.431
DS	0.005	0.007	0.062	0.106	1.000	-0.149
RDS	-0.340	-0.287	-0.057	0.125	-0.055	1.000

^a Firms are ranked for each sample year according to firm size, i.e., equity market value, and assigned into one of three equal-sized groups of firms comprising the small, medium, and large firms.

Variable Definitions:

MVE = market value of common shares outstanding as of fiscal year-end (Compustat #24 * #25);

BVE = book value of common equity as of fiscal year-end (Compustat #60 + #227 - #242);

VCNI = very comprehensive earnings (Compustat #172 - #19 + DS + RDS);

VCN^a = abnormal comprehensive earnings, defined as very comprehensive earnings (VCNI) minus 0.12 * BVE (lagged one year);

DS = dirty surplus, measured as the sum of (1) change in the balance of unrealized gains or losses on marketable securities (change in Compustat #238), (2) change in the cumulative foreign exchange adjustment (change in Compustat #230), and (3) 0.65 times change in additional pension liability in excess of unrecognized prior service costs (change in Min (Compustat #297 - #298, 0)); and

RDS = really dirty surplus, measured as the change in the book value of common equity (Compustat #60 + #227 - #242), less DS, less net income (Compustat #172 - #19), plus dividends (Compustat #21), less share price at middle of fiscal year times change in common shares outstanding (Compustat #25, adjusted for stock dividends and splits).

V. RESULTS

Forecasting Equations

Table 2, Panel A presents regression summary statistics from estimating Equation (4). We employ separate estimations for small, medium, and large firms and the pooled sample. Panel A reveals in all cases, the forecasting coefficient for abnormal very comprehensive income, ω_1 , is significantly positive. It is also increasing in firm size, which is consistent with greater persistence for larger firms.

The incremental forecasting coefficient for *DS*, ω_2 , is significantly different from zero for only the large firms. More importantly, the total *DS* forecasting coefficient, $\omega_1 + \omega_2$, is insignificantly different from zero for all three groups of firms and for the pooled sample (*t*-statistics = 0.46, 0.98, 0.59, and 0.59).¹⁶ These findings indicate that *DS* is forecasting-irrelevant for *VCN*^a for all firms. If investors correctly understand the implications of these persistence findings for valuation, then we should observe valuation irrelevance of *DS* for all firms, i.e., *DS* should have a zero total valuation coefficient in the valuation equation.

The incremental forecasting coefficient for *RDS*, ω_3 , is significantly negative in all cases. The total *RDS* forecasting coefficient, $\omega_1 + \omega_3$, is insignificantly different from zero for all size firms and the pooled sample (*t*-statistics = 1.29, 0.84, -0.85, and -0.81). These findings indicate that *RDS* is forecasting-irrelevant for *VCN*^a for all firms. As with *DS*, if investors correctly understand the implications of these persistence findings for valuation, then we should observe valuation irrelevance of *RDS* for all firms.

Valuation Equations

Table 2, Panel B reveals the valuation coefficient for *VCN*^a, α_1 , is significantly positive in all cases. It is also increasing in firm size, ranging from 1.05 for small firms to 7.83 for large firms, which is consistent with the pattern of increasing persistence displayed in Table 2, Panel A.

The incremental valuation coefficient for *DS*, α_2 , is insignificantly different from zero for small and medium firms and significantly negative for large firms. More importantly, its total coefficient, $\alpha_1 + \alpha_2$, is also insignificantly different from zero for all three groups as well as the pooled sample (*t*-statistics = 1.03, 1.10, -0.39, and -0.43). This finding is expected based on the findings about the lack of persistence for *DS* revealed in Table 2, Panel A.

The incremental valuation coefficient for *RDS*, α_3 , is significantly negative in all cases. Its total coefficient, $\alpha_1 + \alpha_3$, is also significantly negative for all groups (*t*-statistics = -5.45, -8.37, -4.76, and -4.93). Based on the forecasting coefficient findings in Table 2, Panel A, we expect to observe $\alpha_1 + \alpha_3$ to be insignificantly different from zero for all size firms. However, finding that $\alpha_1 + \alpha_3 < 0$ implies that an incremental dollar of *RDS* increases market value by less than a dollar.¹⁷ Thus investors appear to undervalue the *RDS* component of income, i.e., over-value equity.¹⁸

¹⁶ Throughout we use a 0.05 significance level under a two-sided alternative when evaluating statistical significance.

¹⁷ Note that the market's treatment of *RDS* as having negative persistence can be restated as implying that the market views the benefits to the firm from a transaction that give rise to *RDS* as exceeding its *RDS* costs. For this to occur, the market would have to believe that benefits would flow to the firm in future periods at a level beyond that which could be inferred from the time-series properties of residual income. In other words, *RDS* would play two roles—being both a current period cost and a proxy for an Ohlson-type other information variable. An example of this is the market believing an intangible asset arising from employee stock options is greater than the dilution cost to current shareholders.

¹⁸ The findings in Table 2 could be attributable to variables predictive of future earnings and returns that are correlated with *RDS*. Fairfield et al. (2003) identify two such accounting-based variables: short-term accruals and growth in net operating assets. Untabulated findings reveal that inclusion of these variables in Equation (4) and Equation (5) does not affect any inferences we draw from Table 2.

TABLE 2

Regressions of Abnormal Earnings and Equity Market Value, with Dirty Surplus and Really Dirty Surplus Included as Separate Regressors, for a Sample of 37,097 Firm-Year Observations 1976–2006^a

Panel A: Summary Statistics from Regression of Abnormal Earnings on Lagged Abnormal Earnings, Dirty Surplus, Really Dirty Surplus, and Equity Book Value

$$VCNI_{it+1}^a = \omega_0 + \omega_1 VCNI_{it}^a + \omega_2 DS_{it} + \omega_3 RDS_{it} + \omega_4 BVE_{it} + \varepsilon_{it+1}$$

Sample	No. of Obs.	Intercept		VCNI ^a		DS		RDS		BVE		Forecasting Relevance Test	
		Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	$\omega_1 + \omega_2 = 0$	$\omega_1 + \omega_3 = 0$
Small ^b	12,257	0.74	2.88*	0.39	15.91*	-0.28	-1.13	-0.35	-13.15*	-0.07	-10.25*	0.46	1.29
Medium ^b	12,213	4.59	3.99*	0.47	17.85*	-0.15	-0.43	-0.45	-11.73*	-0.04	-5.64*	0.98	0.84
Large ^b	12,627	-16.23	-2.64*	0.53	9.94*	-0.43	-2.35*	-0.59	-6.65*	0.00	0.78	0.59	-0.85
Pooled	37,097	-7.98	-4.10*	0.53	10.07*	-0.43	-2.35*	-0.58	-6.65*	0.00	0.54	0.59	-0.81

Panel B: Summary Statistics from Regression of Market Value of Equity on Abnormal Earnings, Dirty Surplus, Really Dirty Surplus, and Equity Book Value

$$MVE_{it} = \alpha_0 + \alpha_1 VCNI_{it}^a + \alpha_2 DS_{it} + \alpha_3 RDS_{it} + \alpha_4 BVE_{it} + u_{it}$$

Sample	No. of Obs.	Intercept		VCNI ^a		DS		RDS		BVE		Valuation Relevance Test	
		Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.	$\alpha_1 + \alpha_2 = 0$	$\alpha_1 + \alpha_3 = 0$
Small ^b	12,257	8.85	2.17*	1.05	8.66*	1.21	0.55	-3.56	-6.81*	1.23	12.47*	1.03	-5.45*
Medium ^b	12,213	101.80	3.39*	2.72	8.13*	-0.21	-0.09	-5.09	-9.53*	1.51	13.56*	1.10	-8.37*
Large ^b	12,627	766.66	4.11*	7.83	10.37*	-9.30	-2.33*	-10.88	-9.50*	1.93	15.87*	-0.39	-4.76*
Pooled	37,097	187.08	3.27*	7.76	10.39*	-9.37	-2.34*	-10.96	-9.60*	2.01	16.94*	-0.43	-4.93*

* Indicates coefficient is significantly different from 0 at less than the 0.05 level.

^a See Table 1 for definitions of all variables.

^b Firms are ranked for each sample year according to firm size, i.e., equity market value, and assigned into one of three equal-sized groups of firms comprising the small, medium, and large firms.

Dirty Surplus Hedge Returns

Table 3 reports buy-and-hold Fama-French risk-adjusted stock returns for firms in the top and bottom three deciles of firms classified according to the (signed) magnitude of *DS* as a fraction of equity book value at the beginning of the cumulation period. Results are presented separately for small, medium, and large firms, and for the pooled sample. The table presents mean returns for one-, two-, and three-year horizons and the median values of *DS/BVE* for each group, as well as hedge returns and associated t-statistics and empirical p-values.

The findings indicate that for medium and large firms, and for the pooled sample, mean risk-adjusted returns and associated hedge returns for all three horizons are essentially zero. However, for small firms, the mean risk-adjusted returns are positive and monotonically increasing over the three-year horizon, suggesting that Fama-French risk adjustments may not perfectly eliminate the pricing effects of risk. Nonetheless, the small firm hedge returns are zero, indicating that the mismeasurement of expected return for small firms is unrelated to assignment of observations to *DS* portfolios. The small firm hedge returns being zero is consistent with investors pricing *DS* correctly.

Taken together, the findings in Tables 2 and 3 suggest little evidence of mispricing of stocks based on the magnitude of *DS*. Alternatively, the findings suggest investors make no adjustments to reflect *DS* items, but this has no pricing implications because when earnings items are transitory they should only be impounded into price as a result of being included in the book value of equity.

Really Dirty Surplus Hedge Returns

Table 4 reports buy-and-hold Fama-French risk-adjusted stock returns for firms in the top and bottom three deciles of firms classified according to the (signed) magnitude of *RDS* as a fraction of equity book value at the beginning of the cumulation period. Table 4 presents analogous statistics to those presented in Table 3, but are based on *RDS* rather than *DS*. Results are presented separately for small, medium, and large firms, and for the pooled sample.

In contrast to Table 3, the findings in Table 4 generally indicate that risk-adjusted returns differ from zero for all firm groups across all three horizons. In addition, the returns move increasingly away from zero in absolute terms over the investing horizon for almost every portfolio. This is particularly pronounced for small firms, for which the one-, two-, and three-year risk-adjusted returns for the top 30 percent (bottom 30 percent) portfolios are 0.08, 0.12, and 0.17 (0.03, 0.06, 0.08). Recall that the excess returns for small firms in Table 3 also are positive and almost identical for firms in the bottom and top 30 percent portfolios, which we attribute to the difficulty of measuring expected return for small firms. We can therefore treat the excess returns for small firms reported in Table 3 as a benchmark for the measurement error in expected return. Using this approach, we can deduct the average of the two portfolio returns for each investing horizon to arrive at a better estimate of risk-adjusted returns for the small firms. This results in one-, two-, and three-year risk-adjusted returns for the top 30 percent (bottom 30 percent) portfolios of 0.01, 0.02, and 0.02 (−0.06, −0.04, −0.07). Note that these additional risk adjustments for the small firms have no effect on their hedge returns because the same adjustment is made to both the long and short positions.¹⁹

Turning to the hedge returns, Table 4 reveals that they are significantly positive in all three firm-size groups and for the pooled sample. Table 4 also reveals that the hedge returns are

¹⁹ A similar adjustment to excess returns for medium and large firms could be made. However, Table 3 indicates that excess returns for medium and large firms are bounded between −0.01 and 0.02, which suggests that measurement error in expected returns is relatively immaterial for these groups of firms.

TABLE 3
Mean Fama-French Risk-Adjusted Stock Returns in Top and Bottom Three Deciles of
Non-Zero Dirty Surplus Deflated by Book Value of Owner Equity for a Sample of 12,804
Firm-Year Observations
1976–2003^{a,b}

	No. of Obs.	MVE	DS/BVE	1-Year	2-Year	3-Year
Small Firms Portfolio ^c						
Bottom 30%	1,314	162.30	−0.022	0.08	0.11	0.16
Top 30%	1,303	146.09	0.016	0.07	0.10	0.15
Hedge Return ^d				−0.01	0.00	0.00
t-stat.				−0.51	−0.13	−0.07
Empirical p-value ^e				0.72	0.56	0.54
Medium Firms Portfolio ^c						
Bottom 30%	1,279	1,075.51	−0.021	−0.01	0.00	0.01
Top 30%	1,272	934.83	0.016	0.02	0.01	0.00
Hedge Return ^d				0.03	0.00	−0.02
t-stat.				1.71	0.20	−0.62
Empirical p-value ^e				0.03	0.39	0.73
Large Firms Portfolio ^c						
Bottom 30%	1,260	14,073.49	−0.023	0.00	0.00	0.00
Top 30%	1,254	12,525.09	0.016	0.01	−0.01	−0.01
Hedge Return ^d				0.00	−0.01	−0.01
t-stat.				0.34	−0.46	−0.55
Empirical p-value ^e				0.36	0.71	0.73
Pooled Sample Portfolio ^c						
Bottom 30%	3,853	5,014.65	−0.022	0.03	0.04	0.06
Top 30%	3,829	4,462.24	0.016	0.03	0.03	0.05
Hedge Return ^d				0.01	−0.00	−0.01
t-stat.				0.55	−0.19	−0.49
Empirical p-value ^e				0.33	0.58	0.67

^a Indicates hedge return is significantly different from zero at less than the 0.10 level.

^a See Table 1 for definitions of all variables.

^b Fama-French risk-adjusted return is a firm's actual return in excess of the risk-free rate less the firm's predicted return based on the Fama-French factor and momentum factor mimicking portfolios, i.e., excess market return, size, book-to-market, and momentum factor.

^c Firm's size designation and DS portfolio ranking are assigned in the following procedure: for each sample year, firms are ranked according to DS as a fraction of end of year equity book value, BVE, and assigned into ten equal-sized portfolios whereby the first (tenth) portfolio contains those observations with the smallest (largest) fraction of DS; within each of the ten DS portfolios, firms are ranked according to firm size, i.e., equity market value, and are assigned to one of three equal-sized groups of firms comprising the small, medium, and large firms.

^d The hedge return is computed by deducting the mean risk-adjusted return on the bottom three deciles portfolio from that on the top three deciles portfolio. The strategy implementation begins three months subsequent to the firm's fiscal year-end.

^e The proportion of the hedge returns from 1,000 simulations exceeds the observed DS-based hedge return. In a simulation, each firm is assigned a random number as the substitute for DS, and accordingly the portfolio ranking and size designation following the procedure in table note c.

TABLE 4
Mean Fama-French Risk-Adjusted Stock Returns in Top and Bottom Three Deciles of
Really Dirty Surplus Deflated by Book Value of Owner Equity for a Sample of 28,346
Firm-Year Observations
1976–2003^{a,b}

	No. of Obs.	MVE	RDS/BVE	1-Year	2-Year	3-Year
Small Firms Portfolio ^c						
Bottom 30%	2,870	114.23	−0.046	0.03	0.06	0.08
Top 30%	2,860	54.91	0.004	0.08	0.12	0.17
Hedge Return ^d				0.06	0.07	0.09
t-stat.				3.61*	2.61*	2.81*
Empirical p-value ^e				<0.01	0.01	<0.01
Medium Firms Portfolio ^c						
Bottom 30%	2,834	670.28	−0.046	−0.02	−0.05	−0.07
Top 30%	2,824	351.69	0.004	0.01	0.01	0.00
Hedge Return ^d				0.03	0.06	0.07
t-stat.				2.71*	3.92*	3.94*
Empirical p-value ^e				<0.01	<0.01	<0.01
Large Firms Portfolio ^c						
Bottom 30%	2,812	10,156.35	−0.047	−0.02	−0.04	−0.06
Top 30%	2,805	5,625.89	0.004	0.01	0.02	0.02
Hedge Return ^d				0.04	0.06	0.08
t-stat.				4.25*	5.39*	5.83*
Empirical p-value ^e				<0.01	<0.01	<0.01
Pooled Sample Portfolio ^c						
Bottom 30%	8,516	3,615.20	−0.046	−0.01	−0.01	−0.02
Top 30%	8,489	1,994.44	0.004	0.04	0.05	0.07
Hedge Return ^d				0.04	0.06	0.08
t-stat.				5.84*	5.81*	6.00*
Empirical p-value ^e				<0.01	<0.01	<0.01

* Indicates hedge return is significantly different from zero at less than the 0.05 level.

^a See Table 1 for definitions of all variables.

^b Fama-French risk-adjusted return is a firm's actual return in excess of the risk-free rate less the firm's predicted return based on the Fama-French factor and momentum factor mimicking portfolios, i.e., excess market return, size, book-to-market, and momentum factor.

^c Firm's size designation and RDS portfolio ranking are assigned in the following procedure: for each sample year, firms are ranked according to RDS as a fraction of end of year equity book value, BVE, and assigned into ten equal-sized portfolios whereby the first (tenth) portfolio contains those observations with the smallest (largest) fraction of RDS; within each of the ten RDS portfolios, firms are ranked according to firm size, i.e., equity market value, and are assigned to one of three equal-sized groups of firms comprising the small, medium, and large firms.

^d The hedge return is computed by deducting the mean risk-adjusted return on the bottom three deciles portfolio from that on the top three deciles portfolio. The strategy implementation begins three months subsequent to the firm's fiscal year-end.

^e The proportion of the hedge returns from 1,000 simulations exceeds the observed RDS-based hedge return. In a simulation, each firm is assigned a random number as the substitute for RDS, and accordingly the portfolio ranking and size designation following the procedure in table note c.



increasing over time. For example, for small firms, the one-, two-, and three-year hedge returns are 0.06, 0.07, and 0.09. It is possible that the increasing hedge returns over time could be attributable to an unmodeled risk factor.^{20,21}

Additional RDS Hedge Return Tests

In this section, we consider several additional tests to examine the sensitivity of the *RDS* hedge returns to previously documented pricing anomalies and risk factors, and to alternative procedures for computing those returns. We also attempt to determine the extent to which different source components of *RDS* account for our hedge return results.

First, we investigate whether the persistence of the *RDS* hedge returns could reflect the effects of pricing anomalies documented in prior research. We consider each anomaly, in turn, by first placing observations into one of ten portfolios based on the magnitude of each anomaly factor. Then, within each portfolio, we rank observations according to the magnitude of *RDS* as a fraction of *BVE*, assigning each observation within each anomaly portfolio to one of ten *RDS* portfolios. This double-sorting process helps to ensure that our findings are not the result of the particular anomaly used in the initial sort.

The pricing anomalies we consider are: the short-term accruals anomaly (Sloan 1996; Xie 2001); the growth in long-term asset accrual anomaly (Fairfield et al. 2003); the long-term pricing reversal anomaly (Daniel and Titman 2006; Fama and French 2008a), and the share repurchase and issuance anomaly (Ikenberry et al. 1995; Daniel and Titman 2006; Mitchell and Stafford 2000; Fama and French 2008a). The share repurchase and issuance anomaly is potentially most closely related to the *RDS* pricing anomaly we document because the latter can only arise from equity transactions. Untabulated findings reveal that *RDS* hedge returns remain significantly positive after controlling for the potential confounding effects of each of the four anomalies. Thus, mispricing associated with previously documented anomalies appears not to account for our finding investor mispricing of *RDS*.

Second, we consider an alternative to our buy-and-hold procedure for computing *RDS* hedge returns. Following Mashruwala et al. (2006), we estimate hedge returns using a monthly calendar-time portfolio approach (Fama and French 1993). Under this approach, hedge returns are calculated based on Jensen's alphas from monthly time-series regressions of hedge portfolio excess returns on Fama-French and momentum factor returns. For each sample year, we assign firms into decile portfolios based on *RDS* as a fraction of equity book value and compute the mean monthly portfolio return for firms in the top and bottom three deciles. We then estimate the following monthly time-series regression for both the high and low *RDS* portfolios:

$$R_{p,t} - R_{f,t} = \alpha_p + \beta_p(R_{M,t} - R_{f,t}) + s_pSMB_t + h_pHML_t + m_pMOM_t + \varepsilon_{p,t}. \quad (8)$$

The resulting Jensen's alpha, α_p , measures the mean monthly return for portfolio *p* not attributable to the Fama-French and momentum factor returns. We predict that α_p for the high *RDS* portfolio is larger than that for the low *RDS* portfolio. We formally test this by estimating Equation (8) for a hedge portfolio, which is constructed as the difference in mean monthly excess returns for the

²⁰ One possible candidate is firm size, as Table 4 indicates that firms in the bottom 30 percent *RDS* portfolios are roughly twice as large as those in the top 30 percent. To assess the importance of firm size on hedge returns, we regressed excess return on firm size and an indicator variable for whether a firm-year observation is in the top 30 percent *RDS* portfolio. Untabulated findings indicate that the indicator variable coefficient is significantly positive in all cases and over all horizons, and the coefficient on size is insignificant in all specifications.

²¹ In addition, it is possible that significant hedge returns are induced by our implementation of risk-adjusting returns. Untabulated hedge returns computed without explicit adjustment for risk yield inferences are consistent with those based on the tabulated findings.

high *RDS* and low *RDS* portfolios,²² and then testing whether the resulting Jensen's alpha is significantly positive.

Consistent with our predictions, untabulated findings indicate that Jensen's alpha for the high *RDS* portfolio is larger than that for the low portfolio for all three firm-size groups: 0.005 versus 0.003 (small firms); 0.001 versus 0.000 (medium firms); 0.001 versus 0.000 (large firms). For the pooled sample, the high and low portfolio Jensen's alpha are 0.002 and 0.000, respectively. The Jensen's alpha of the hedge portfolio is statistically significant for only the small firms and the pooled sample, 0.003 (t-statistic = 2.25; p-value = 0.03) and 0.002 (t-statistic = 2.27; p-value = 0.02), which in annual terms indicate excess returns to the *RDS*-based hedging strategy of 3.7 percent and 2.4 percent, respectively. These findings contrast with those from the buy-and-hold hedge returns in Table 4, which indicate hedge returns are positive for all three firm-size groups. Ascertaining which approach yields the more reliable results is not straightforward. For example, the buy-and-hold approach has the advantage of updating the individual stock's expected return on a monthly basis using out-of-sample estimation, and the alpha approach assumes that factor betas are constant during the test period.²³

Third, we attempt to determine the extent to which different source components of *RDS* account for our hedge return results. Findings from the forecasting and valuation tests are consistent with, in the aggregate, *RDS* components of very comprehensive net income being transitory, but investors failing to understand this and over-valuing equity. Ideally, we would like to identify separate components of *RDS*, determine their persistence, and then ascertain which components investors appear to fail to price correctly. Because we face the same data limitations as investors, we can only do this indirectly by sequentially excluding and including firm-years in which *RDS* is more likely to be attributable to one particular type of transaction. First, to focus on the potential pricing effects of pooling-of-interests transactions, we recomputed hedge returns (1) excluding firm-years with corporate mergers or acquisitions, and (2) using only firm-years with corporate mergers.²⁴ Second, to focus on the potential pricing effects of employee stock options, we recomputed hedge returns using (1) only observations beginning in 1995, and (2) only observations before 1995. Because SFAS No. 123 required disclosure of weighted average exercise price for employee stock options after 1995, hedge returns might be expected to fall or even disappear in this latter period. Third, to focus more generally on the pricing effects of dilutive transactions, including warrants, convertible instruments, as well as stock options, we recomputed hedge returns for portfolios sorting firm-year observations on the difference between basic and diluted earnings per share as a fraction of equity book value. Finally, because *RDS* is concentrated in particular industries and years, as noted in Section IV, we sequentially calculated hedge returns excluding the six industries and the five years with the most negative *RDS*.

Untabulated findings from each of these additional analyses reveal that hedge returns remain significant in all cases. These findings are consistent with our tests lacking power to trace the precise sources of *RDS*. They are also consistent with *RDS* being correlated with an unmodeled risk factor.

²² The independent variables (risk factors) are the same for each portfolio and are therefore included in the hedge portfolio without adjustment.

²³ For a discussion of the strengths and weaknesses of the various approaches, see Fama and French (2008b).

²⁴ Ideally, we would compute hedge returns for subsamples excluding and including only those firm-years with mergers accounted for under pooling-of-interests. However, because such identifying information is only available to us using the Securities Data Corporation beginning in the middle of our sample period, we elected to cast the net wider to take account of mergers and acquisitions during our sample period.

VI. CONCLUSIONS

The question we address in this study is whether firms' share prices correctly reflect two accounting measures, dirty surplus and really dirty surplus. We find that both dirty surplus and really dirty surplus are forecasting-irrelevant for abnormal very comprehensive income for all firm-size groups. Taking these results at face value (i.e., assuming that the forecasting and valuation equations correctly capture the time-series properties of these variables), if investors correctly understand the implications of these persistence findings for valuation, then each kind of dirty surplus should be valuation-irrelevant for all firms. This prediction is borne out in the case of dirty surplus. In contrast, the findings indicate that investors appear to undervalue really dirty surplus, which is consistent with the premise that investors are unable to assess the economic implications of really dirty surplus transactions. However, the possibility remains that the mismatch of the really dirty surplus forecasting and valuation coefficients is the result of model misspecification rather than mispricing.

Our buy-and-hold hedge return results support the findings from the tests linking the forecasting and valuation equations. Hedge returns are insignificantly different from zero when based upon dirty surplus, regardless of firm size and investing horizon. In contrast, buy-and-hold hedge returns based on really dirty surplus are significantly positive for all three firm-size groups as well as for the pooled sample. Findings from additional tests reveal that inferences relating to hedge returns are insensitive to an alternative procedure to measuring hedge returns, to including controls for four previously identified mispricing anomalies, and to sampling procedures designed to focus on sources of really dirty surplus.

Taken together, the findings are consistent with investors failing to understand the lack of persistence of really dirty surplus, and therefore apparently over-valuing firms that have large negative really dirty surplus. However, because we are unable to trace the sources of really dirty surplus to particular types of equity transactions, we cannot determine the extent to which potential mispricing arises from each type of transaction. However, even if we could trace the sources of really dirty surplus, any resulting hedge returns might still be attributable to an unmodeled risk factor.

REFERENCES

- Ang, A., and J. Liu. 2004. How to discount cash flows with time-varying expected returns. *The Journal of Finance* 59 (6): 2745–2783.
- Barth, M. E., and S. Kallapur. 1996. Effects of cross-sectional scale differences on regression results in empirical accounting research. *Contemporary Accounting Research* 13 (2): 527–567.
- , W. H. Beaver, J. M. Hand, and W. R. Landsman. 1999. Accruals, cash flows, and equity values. *Review of Accounting Studies* 4 (3–4): 205–229.
- , ———, ———, and ———. 2005. Accruals, accounting-based valuation models, and the prediction of equity values. *Journal of Accounting, Auditing & Finance* 20 (4): 311–345.
- , Y. Konchitchki, and W. R. Landsman. 2008. Cost of capital and financial statements transparency. Working paper, Stanford University.
- Bell, T., W. R. Landsman, B. Miller, and S. Yeh. 2002. The valuation implications of employee stock option accounting for profitable computer software firms. *The Accounting Review* 77 (4): 971–996.
- Bernard, V., and J. Thomas. 1989. Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research* 27 (Supplement): 1–48.
- , and ———. 1990. Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics* 13 (4): 305–340.
- Biddle, G., and J. Choi. 2006. Is comprehensive income useful? *Journal of Contemporary Accounting and Economics* 2 (1): 1–32.
- Bradshaw, M., and R. G. Sloan. 2002. GAAP versus the Street: An empirical assessment of two alternative

- definitions of earnings. *Journal of Accounting Research* 40 (1): 41–66.
- Brown, L., and K. Sivakumar. 2003. Comparing the value relevance of two operating income measures. *Review of Accounting Studies* 8 (4): 561–572.
- Burgstahler, D., J. Jiambalvo, and T. Shevlin. 2002. Do stock prices fully reflect the implications of special items for future earnings? *Journal of Accounting Research* 40 (3): 585–612.
- Carhart, M. 1997. On persistence in mutual fund performance. *The Journal of Finance* 52 (1): 57–82.
- Chambers, D., T. J. Linsmeier, C. Shakespeare, and T. Sougiannis. 2007. An evaluation of SFAS No. 130 comprehensive income disclosures. *Review of Accounting Studies* 12 (4): 557–593.
- Christensen, P., and G. A. Feltham. 2003. *Economics of Accounting*. Boston, MA: Kluwer Academic Publishers.
- Core, J., W. Guay, and S. P. Kothari. 2002. The economic dilution of employee stock options: Diluted EPS for valuation and financial reporting. *The Accounting Review* 77 (3): 627–652.
- Daniel, K., and S. Titman. 2006. Market reactions to tangible and intangible information. *The Journal of Finance* 61 (4): 1605–1643.
- Dechow, P. M., A. P. Hutton, and R. G. Sloan. 1999. An empirical assessment of the residual income valuation model. *Journal of Accounting and Economics* 26 (1–3): 1–34.
- Dhaliwal, D., K. R. Subramanyam, and R. Trezevant. 1999. Is comprehensive income superior to net income as a measure of firm performance? *Journal of Accounting and Economics* 26 (1–3): 43–67.
- Doyle, J. T., R. J. Lundholm, and M. T. Soliman. 2003. The predictive value of expenses excluded from ‘pro forma’ earnings. *Review of Accounting Studies* 8 (2–3): 145–174.
- Fairfield, P. M., J. S. Whisenant, and T. L. Yohn. 2003. Accrued earnings and growth: Implications for future profitability and market mispricing. *The Accounting Review* 78 (1): 353–371.
- Fama, E. F., and K. R. French. 1993. Common-size risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33 (1): 3–56.
- , and ———. 2008a. Average returns, B/M, and share issues. *The Journal of Finance* 63 (6): 2971–2995.
- , and ———. 2008b. Dissecting anomalies. *The Journal of Finance* 63 (4): 1653–1678.
- Feltham, G. 1996. Valuation, clean surplus accounting, and anticipated equity transactions. Working paper, University of British Columbia.
- , and J. Ohlson. 1995. Valuation and clean surplus accounting for operating and financial activities. *Contemporary Accounting Research* 11 (2): 689–731.
- Gow, I. D., G. Ormazabal, and D. J. Taylor. 2010. Correcting for cross-sectional and time-series dependence in accounting research. *The Accounting Review* 85 (2): 483–512.
- Ibbotson, and Associates. 2005. *Cost of Capital 2005 Yearbook*. Chicago, IL: Ibbotson and Associates.
- Ikenberry, D., J. Lakonishok, and T. Vermaelen. 1995. Market underreaction to open market share repurchases. *Journal of Financial Economics* 39 (2–3): 181–208.
- Jegadeesh, N., and S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance* 48 (1): 65–91.
- Landsman, W. R., B. L. Miller, and S. Yeh. 2007. Implications of components of income excluded from pro forma earnings for future profitability and equity valuation. *Journal of Business Finance & Accounting* 34 (3–4): 650–675.
- , K. Peasnell, P. Pope, and S. Yeh. 2006. Which approach to accounting for employee stock options best reflects market pricing? *Review of Accounting Studies* 11 (2–3): 203–245.
- Mashruwala, C., S. Rajgopal, and T. Shevlin. 2006. Why is the accrual anomaly not arbitrated away? The role of idiosyncratic risk and transactions costs. *Journal of Accounting and Economics* 42 (1–2): 3–33.
- Massa, M., U. Peyer, and Z. Tong. 2005. Limits of arbitrage and corporate financial policy. Working paper, INSEAD.
- Mitchell, M. L., and E. Stafford. 2000. Managerial decisions and long-term stock price performance. *The Journal of Business* 73 (3): 287–329.
- O’Hanlon, J., and P. Pope. 1999. The value-relevance of U.K. dirty surplus accounting flows. *The British Accounting Review* 31 (4): 459–482.
- Ohlson, J. A. 1995. Earnings, equity book values, and dividends in equity valuation. *Contemporary Accounting Research* 11 (2): 661–687.

- . 1999. On transitory earnings. *Review of Accounting Studies* 4 (3–4): 145–162.
- Petersen, M. A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22 (1): 435–480.
- Sloan, R. 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71 (3): 289–316.
- Xie, H. 2001. The mispricing of abnormal accruals. *The Accounting Review* 76 (3): 357–373.

Can Big 4 versus Non-Big 4 Differences in Audit-Quality Proxies Be Attributed to Client Characteristics?

Alastair Lawrence
Miguel Minutti-Meza
Ping Zhang
University of Toronto

ABSTRACT: This study examines whether differences in proxies for audit quality between Big 4 and non-Big 4 audit firms could be a reflection of their respective clients' characteristics. In our analyses, we use three audit-quality proxies—discretionary accruals, the *ex ante* cost-of-equity capital, and analyst forecast accuracy—and employ propensity-score and attribute-based matching models in attempt to control for differences in client characteristics between the two auditor groups while estimating the audit-quality effects. Using these matching models, we find that the effects of Big 4 auditors are insignificantly different from those of non-Big 4 auditors with respect to the three audit-quality proxies. Our results suggest that differences in these proxies between Big 4 and non-Big 4 auditors largely reflect client characteristics and, more specifically, client size. We caution the reader that this study has not resolved the question, although we hope that it encourages other researchers to explore alternative methodologies that separate client characteristics from audit-quality effects.

Keywords: *Big 4 versus non-Big 4 audit quality; discretionary accruals; ex ante cost-of-equity capital; analyst forecast accuracy; propensity-score matching; attribute-based matching.*

Data Availability: *Data are publicly available from sources identified in the article.*

We have received valuable comments from Steven Kachelmeier (senior editor), Francesco Bova, Gus DeFranco, Ole-Kristian Hope, Stephannie Larocque, Hai Lu, Flora Niu, Gordon Richardson, Kevin Veenstra, Dushyantkumar Vyas, two anonymous reviewers, and workshop participants at the University of Toronto and Wilfrid Laurier University. Professor Zhang acknowledges the financial support provided by the Social Sciences and Humanities Research Council of Canada.

Editor's note: Accepted by Steven Kachelmeier.

Submitted: June 2008
Accepted: June 2010
Published Online: January 2011

I. INTRODUCTION

The auditing literature generally concludes that the audit quality of Big 4 auditors is superior to that of non-Big 4 auditors.¹ DeAngelo (1981) argues that accounting firm size is a proxy for auditor quality, as no single client is important to larger accounting firms and, hence, larger accounting firms are less likely than smaller accounting firms to compromise their independence. Dopuch and Simunic (1980) propose that larger accounting firms provide higher quality services because they have greater reputations to protect. Furthermore, it could be argued that Big 4 firms provide superior audit quality as their sheer size can support more robust training programs, standardized audit methodologies, and more options for appropriate second partner reviews.

However, there are also arguments as to why Big 4 and non-Big 4 firms could provide comparable audit quality. First, Big 4 and non-Big 4 firms are held to the same regulatory and professional standards, and thus both types of audit firms must adhere to a reasonable level of quality. Second, as “non-Big 4 auditors have superior knowledge of local markets and better relation with their clients” (Louis 2005, 77), these factors may enable non-Big 4 firms to better detect irregularities. Of course, the converse argument could be made that closer relationships among non-Big 4 accounting firms and their clients could potentially lead to a compromise of independence; however, the net effect of these counteracting forces is unclear. Third, the inability of non-Big 4 firms to obtain affordable insurance coverage may actually increase the audit effort of non-Big 4 firms relative to Big 4 firms because smaller audit firms cannot obtain a similar level of backing from insurance companies. This notion is supported by the U.S. Government Accountability Office (GAO) report issued in 2008 indicating that non-Big 4 auditors are struggling to obtain affordable liability insurance coverage (GAO 2008, 55). Moreover, the previous GAO audit firm concentration report issued in 2003 finds that “smaller firms lacked the size needed to achieve economies of scale to spread their litigation risk and insurance costs across a larger capital base” (GAO 2003, 49), suggesting that insurance fees are a fixed cost to audit partners. Finally, CPAs frequently switch between Big 4 and non-Big 4 firms, and knowledge transfers can dilute the potential for one type of audit firm to become superior. For example, upon the fallout of Arthur Andersen (AA), Grant Thornton acquired AA’s Albuquerque, Charlotte, Columbia, Greensboro, and Orlando offices and separately hired many former middle-market AA audit partners from other offices (Dow Jones News Service 2002). Hence, it is not obvious from theory or intuition that Big 4 firms should be superior to non-Big 4 firms.

Nonetheless, numerous empirical studies, following the theoretical support of DeAngelo (1981) and Dopuch and Simunic (1980), and using a variety of audit-quality proxies, find evidence suggesting that Big 4 auditors provide higher-quality audits than non-Big 4 auditors (e.g., Palmrose 1988; Becker et al. 1998; Khurana and Raman 2004; Behn et al. 2008). Given that the distributions of client characteristics significantly differ across Big 4 and non-Big 4 firms, an important consideration of these prior studies is whether the empirical findings simply reflect client and not accounting firm characteristics. Our research objective is to investigate whether the Big 4 treatment effect may be attributable to client characteristics.

The question of Big 4 superiority is important, given that many studies rely on the Big 4 versus non-Big 4 distinction as an audit-quality proxy. Hence, it is prudent to confirm that this distinction does not simply reflect client characteristics.² Furthermore, incorrectly classifying Big

¹ We use Big 4 as a generic term encompassing the Big 8, Big 6, Big 5, and Big 4 to reflect the consolidation of these firms.

² Several studies outside the auditing literature have used the Big 4 distinction to proxy for a higher level of audit quality; for example, Beatty (1989), Guenther and Willenborg (1999), Mitton (2002), Smart and Zutter (2003), and Gul et al. (2009).

4 auditors as superior to non-Big 4 auditors has unnecessary negative ramifications for smaller auditors, such as audit committee's auditor selection bias (CFA Institute Center 2009) and discriminatory clauses in loan and underwriting agreements (DeAngelo 1981), which could result in a loss of current and future clients.

Using three audit-quality proxies—discretionary accruals, the *ex ante* cost-of-equity capital, and analyst forecast accuracy—we replicate previous findings that document differences in quality between Big 4 and non-Big 4 audit firms. Following Rosenbaum and Rubin (1983), Li and Prabhala (2007), and Francis et al. (2010), we use propensity-score matching models in an attempt to control for differences in client characteristics between the two auditor groups while estimating auditor treatment effects. Moreover, we match on specific client characteristics (attribute-based matching), such as client size and return on assets, to identify whether Big 4 versus non-Big 4 differences in audit-quality proxies can be attributed to specific client characteristics. Using these matching models, we find that the treatment effects of Big 4 auditors are insignificantly different from those of non-Big 4 auditors with respect to our three audit-quality proxies. Furthermore, the attribute-matching results provide evidence for the argument that differences in the foregoing proxies between Big 4 and non-Big 4 clients largely reflect client size. Using our full samples, we control for an extensive list of firm characteristics known to influence audit quality and find that the Big 4 effect is generally insignificant, indirectly supporting the argument that the Big 4 distinction may reflect client and not auditor characteristics.

Our results must be interpreted with due regard to their limitations and to the caveats of matching models, discussed in more detail throughout the article. This study has not resolved the underlying question as to whether differences in audit-quality proxies between Big 4 and non-Big 4 auditors can be attributed to client characteristics, but we hope that it encourages other researchers to explore alternative methodologies that further disentangle client characteristics from audit-quality effects.

II. RELATED STUDIES AND MEASURES OF AUDIT QUALITY

As the main observable outcome of an audit is the standardized audit report, researchers have used various proxies in an attempt to assess audit quality and, in turn, determine whether a differential in audit quality exists. An extensive branch of audit differentiation research focuses on the quality of the client's financial statements, in which discretionary accruals are often used as a proxy for audit quality, as they reflect the auditor's constraint over management's reporting decisions. Becker et al. (1998) find that Big 4 clients report lower absolute discretionary accruals than non-Big 4 clients. Francis et al. (1999) suggest that Big 4 auditors constrain opportunistic and aggressive reporting because their clients have higher total accruals but lower discretionary accruals. Krishnan (2003) finds a greater association between discretionary accruals and future earnings for Big 4 than for non-Big 4 clients. Following this literature, we use discretionary accruals as our first measure of audit quality. The benefit of this measure is that it reflects the auditor's enforcement of accounting standards. However, a weakness is that it only partially captures the effectiveness of an audit in constraining earnings management, as discretionary accruals not only reflect management's opportunism, but also management's signaling attempts and random noise, as noted by Guay et al. (1996).

The valuation literature suggests that Big 4 auditors provide more assurance to the market than non-Big 4 auditors. The underlying intuition for using valuation proxies as a measure of audit quality is due to the fact that if market participants perceive that the Big 4 clients have more credible earnings than those of the non-Big 4 clients then, *ceteris paribus*, the Big 4 clients should receive a break in their cost-of-equity capital. For example, Khurana and Raman (2004) find that Big 4 clients have a lower *ex ante* cost of capital than non-Big 4 clients in the U.S.; however, they do not find such a difference in Australia, Canada, or Great Britain. Following Khurana and

Raman (2004), we use the *ex ante* cost-of-equity capital as our second audit-quality proxy to capture the capital market's perception of the financial reporting credibility of Big 4 and non-Big 4 clients.

More recently, Behn et al. (2008) include analyst forecast accuracy as an audit-quality proxy. They argue that if one type of auditor increases the reporting reliability of earnings in comparison to the other type, then, *ceteris paribus*, analysts of the superior type's clients should be able to make more accurate forecasts of future earnings than those analysts of the non-superior type's clients. Using this reasoning, Behn et al. (2008) find that analysts of Big 4 clients have higher forecast accuracy than analysts of non-Big 4 clients. We use analyst forecast accuracy as our third audit-quality measure to proxy for an enhanced level of decision making by sophisticated financial statement users.

In summary, we employ three audit-quality proxies—discretionary accruals, the *ex ante* cost-of-equity capital, and analyst forecast accuracy—to capture various aspects of audit quality in comparing Big 4 to non-Big 4 firms.³

III. MATCHING PROCEDURES AND TESTS OF NONLINEARITY

We use propensity-score matching models, developed by Rosenbaum and Rubin (1983), to match on a broad range of client characteristics and use attribute-based matching to examine whether the Big 4 distinction can be attributed to specific client characteristics. Propensity-score matching models match observations based on the probability of undergoing the treatment, which in our case is the probability of selecting a Big 4 auditor. Matching models have important features appropriate for our setting. First, these models generate samples in which the clients of Big 4 and non-Big 4 auditors are similar, providing a natural framework to parse out the effects of auditor and client characteristics on the audit-quality proxy. Second, previous studies examining the existence of a differential in audit quality typically use Heckman (1979) selection models that rely on a specific functional form to provide an indirect estimate of auditor treatment effects. Matching models do not rely on a specific functional form and provide a more direct estimate of the treatment effects (Li and Prabhala 2007). Moreover, selection models pertaining to auditor choice could estimate treatment effects with error by failing to meet the exclusion restrictions because it is difficult to identify variables that influence the auditor selection (first-stage regression) and not the audit-quality proxy (second-stage regression). Last, matching models mitigate the potential impact of nonlinearities in estimating the treatment effects when the underlying functional form is nonlinear. A recent example of a study using a propensity-score matching model in the accounting literature is Armstrong et al. (2010), who find that the previously documented relation between equity-based compensation and accounting irregularities does not hold using propensity-score matching models.

Notwithstanding their benefits, matching models have caveats. First, matching models rely on the assumption that the effects of unobservables are not pertinent to estimating treatment effects. Second, matching results in using subsamples of the population, and hence, as noted by Cram et al. (2009), matching reflects a trade-off between identifying the treatment effects and generalizing the results to the full population. Third, matching results in a different auditor composition in the

³ Other branches of the auditor differentiation literature use litigation, going concern, and fraud frequencies as audit-quality proxies. These branches suggest that Big 4 firms are sued and sanctioned less (Palmrose 1988; Feroz et al. 1991), have lower thresholds for and greater accuracy in issuing opinions (Francis and Krishnan 1999; Lennox 1999; Weber and Willenborg 2003; Geiger and Rama 2006), and have a lower incidence of fraud among their clients than non-Big 4 firms (Farber 2005). We do not use these proxies because the infrequent occurrence of these events results in small sample sizes with low power, potentially biasing in favor of our inferences. For example, Dechow et al. (1996) and Beneish (1999), using small samples, do not find a significant difference between the fraud incidence of Big 4 and non-Big 4 clients.

matched samples than in the full population. These features of matching models could bias our findings due to the reduced power of the tests or due to systematic differences in the subsamples from the full population. Fourth, the empirical inability to match on pre-treatment attributes and control for alternative treatments could result in a bias if the matching variables are affected by the auditor choice. Despite our attempts, we cannot fully rule out concerns that our inferences are affected by *ex post* matching or alternative treatments. In the “Sensitivity Analyses” section, we document a series of robustness tests in attempt to mitigate some of these concerns.

We use a logit model to estimate the probability of selecting a Big 4 auditor, as it is the most prevailing approach for estimating propensity scores (Guo and Fraser 2010, 135).⁴ Given that matching models do not require exclusion restrictions, the general rule is to include a comprehensive list of attributes when estimating the propensity score (Li and Prabhala 2007). We estimate the propensity score for each audit-quality proxy analysis, including variables from the selection model used in Chaney et al. (2004) and the respective audit quality regression, as follows:⁵

$$BIG4_{i,t} = \beta_0 + \beta_1 LOG\_ASSETS_{i,t} + \beta_2 ATURN_{i,t} + \beta_3 CURR_{i,t} + \beta_4 LEV_{i,t} + \beta_5 ROA_{i,t} + \Sigma PROXY\_CONTROLS_{i,t} + Industry\_FE + Year\_FE + \varepsilon_{i,t}, \quad (1)$$

where for firm i and fiscal year t :

$BIG4 = 1$ if the client has a Big 4 auditor in year $_t$, and 0 otherwise;
 LOG_ASSETS = natural logarithm of total assets at the end of year $_t$;
 $ATURN$ = sales/total assets $_{t-1}$;
 $CURR$ = current assets/current liabilities $_t$;
 LEV = (long term debt $_t$ plus debt in current liabilities $_t$)/average total assets $_{t-1}$;
 ROA = net income $_t$ /average total assets $_{t-1}$; and,
 $PROXY\_CONTROLS$ = control variables used and described in the respective audit-quality analysis.

We estimate the propensity-score model by including all audit proxy control variables for the respective analysis, but we exclude variables from the regression if they are redundant to the first five variables in predicting auditor choice.⁶ We then match, without replacement, a non-Big 4 audit client with a Big 4 audit client that has the closest predicted value from Equation (1) within a maximum distance of 3 percent.⁷ Using this caliper distance, we match approximately 80 percent, 99 percent, and 97 percent of non-Big 4 audit clients to Big 4 audit clients for the discretionary accruals, the *ex ante* cost-of-equity capital, and the analyst forecast-accuracy samples, respectively. In effect, this procedure creates a pseudo “random” sample in which the auditor type is randomly allocated to both the treatment and control groups (Heckman and Navarro-Lozano 2004), such that any resulting differences between the two groups should reflect the treatment

⁴ All findings documented in this study are robust to using a probit model instead of a logit model to calculate propensity scores. In addition, we find that the logit models result in more balanced matched samples than using the probit models.

⁵ We also estimate and match the propensity scores within each industry and year rather than using indicator variables. We find that all results using this alternative specification are similar to the tabulated results.

⁶ Alternatively, we estimate the propensity-score model by including all the variables in the model as specified above even when a proxy control variable is similar to a variable in the auditor selection model. We find that the results are robust to including all the redundant variables simultaneously or to including only one redundant variable at a time.

⁷ Inferences are the same whether we match with or without replacement. A caliper is the distance between the predicted probabilities of choosing the treatment between matched observations (Dehejia and Wahba 2002).

effect and not pre-existing client characteristics (Heckman et al. 1997, 1998; Dehejia and Wahba 1999, 2002). Hence, differences in means between the treatment and control groups should be sufficient to estimate the treatment effects (Dehejia and Wahba 2002; Dehejia 2005). Nonetheless, we also use multivariate analyses to further control for any remaining characteristic imbalances between the two client groups as well as general cross-sectional characteristic variations.

We employ the Ramsey Regression Equation Specification Error Test (Ramsey 1969; hereafter, RESET test), to identify specific client characteristics that are nonlinear to both the auditor choice and the audit-quality proxy.⁸ For each propensity-score regression and each audit-quality regression, we add squared and cubed versions of all control variables, as described in Sections V–VII. Using an F-test, we test the null hypothesis that all of the nonlinear terms of each control variable equal zero. If significant nonlinearities exist in the data and the Big 4 partitioning variable is correlated with these nonlinear variables, then these conditions could potentially confound the inferences of a standard OLS regression. For the attribute-based matching, we create single-variable matched samples in each analysis, using the propensity-score methodology controlling for year and industry effects, based on those variables that are nonlinear to both the auditor choice and the audit-quality proxy.

IV. SAMPLE SELECTION

For our analyses, we use firm-year data from 1988 to 2006. We restrict our analyses to this time period because reported operating cash flows as per SFAS No. 95 (FASB 1987) are only available starting from 1988. For our discretionary accruals sample, we use data from Compustat. After deleting firms in the financial services industries (SIC codes between 6000–6999) and after imposing all the necessary requirements to calculate the discretionary accruals regression variables, we obtain a sample of 72,600 firm-year observations, in which 59,323 and 13,277 are Big 4 and non-Big 4 clients, respectively.⁹ Our *ex ante* cost-of-equity sample reflects the intersection of Compustat, I/B/E/S Summary, and CRSP data. After imposing the necessary requirements to calculate the *ex ante* cost-of-capital regression variables, we obtain a sample of 25,068 firm-year observations, in which 23,856 and 1,212 reflect Big 4 and non-Big 4 clients, respectively.¹⁰ Similar to the *ex ante* cost-of-equity sample, the forecast-accuracy sample reflects the intersection of Compustat, I/B/E/S, and CRSP data; however, the accuracy sample requires individual analyst forecast data from the I/B/E/S Detail files, whereas the *ex ante* cost-of-equity sample requires consensus forecast data from the I/B/E/S Summary files. After imposing the necessary requirements to calculate the forecast-accuracy regression variables, we obtain a sample of 28,037 firm-year observations, in which 26,521 and 1,516 reflect Big 4 and non-Big 4 clients, respectively.

⁸ As in predicting the propensity scores, we exclude variables from the RESET analysis if the two regressions contain redundant variables. The comparison of the RESET findings across the three audit-quality proxies is inherently limited given that we adopt the audit-quality regression models for each proxy from previous studies; thus, there are limited control variables that are common to all three analyses.

⁹ We delete firms with negative total assets, sales, debt, and market values of equity, as such observations introduce noise into the discretionary accruals regressions; moreover, raw variables from Compustat are winsorized at the 1 percent and 99 percent levels. We find that inferences for our three analyses hold without winsorizing except for those pertaining to the propensity-score matched sample in the analyst forecast accuracy analysis, as significant outliers reduce the effectiveness of the multivariate propensity-score matching.

¹⁰ For the *ex ante* cost-of-equity, and forecast-accuracy samples, we delete firms in the financial services industries (SIC codes between 6000–6999) and all variables are winsorized at the 1 percent and 99 percent levels.

V. ANALYSIS 1: DISCRETIONARY ACCRUALS

Method

We measure performance-adjusted discretionary accruals using the modified Jones model as recommended by Kothari et al. (2005; hereafter K LW).¹¹ To test whether the Big 4 versus non-Big 4 discretionary accruals differences could be attributed to client characteristics, we use the following model:

$$ADA_{i,t} = \beta_0 + \beta_1 BIG4_{i,t} + \beta_2 LOG\_MKT_{i,t} + \beta_3 ROA_{i,t-1} + \beta_4 LEV_{i,t-1} + \beta_5 CURR_{i,t-1} + Industry\_FE + Year\_FE + \varepsilon_{i,t}, \quad (2)$$

where for firm i and fiscal year t :

ADA = absolute discretionary accruals as per K LW in year;
 $BIG4$ = 1 if the client has a Big 4 auditor in year, and 0 otherwise;
 LOG_MKT = natural logarithm of market value of equity at the end of year;
 ROA = income before extraordinary items _{$t-1$} /average total assets _{$t-1$} ;
 LEV = (long-term debt _{$t-1$} plus debt in current liabilities _{$t-1$})/average total assets _{$t-1$} ; and
 $CURR$ = current assets _{$t-1$} /current liabilities _{$t-1$} .

$BIG4$ is our main variable of interest and is used in a manner consistent with prior research. We include the variable LOG_MKT to control for client size. Following Butler et al. (2004), we include ROA , as the K LW procedure does not eliminate all the variation in accruals due to firm-specific performance. Also in line with Butler et al. (2004), we include LEV and $CURR$ to control for the impact of financial risk on discretionary accruals.

Results

Table 1 presents the descriptive statistics for both the full and propensity-score matched samples. There are 72,600 firm-year observations in the full sample, of which 59,323 (77.6 percent) and 13,277 (22.4 percent) reflect Big 4 and non-Big 4 clients, respectively. The descriptive statistics for the full sample indicate that Big 4 and non-Big 4 auditors have significantly different clienteles. Big 4 clients are approximately ten times the size of non-Big 4 clients: the mean market value and total assets of Big 4 clients are \$2.2 billion and \$2.0 billion, respectively, whereas the mean market value and total assets of non-Big 4 auditors are \$234 million and \$258 million, respectively. Moreover, Big 4 clients are more profitable and leveraged than are non-Big 4 clients, and have significantly less discretionary accruals and current assets than non-Big 4 clients.

Using Equation (1) to calculate the propensity scores and imposing a caliper distance of 3 percent, we obtain a propensity-score matched sample of 21,130 firm-years, of which 10,565 are Big 4 clients and 10,565 are non-Big 4 clients. The propensity-score model appears effective in forming a balanced sample of Big 4 and non-Big 4 clients, as all accrual control variables in the

¹¹ The K LW model is as follows and is estimated by year and by two-digit SIC code, scaling by average lagged assets:

$$TOTAL\_ACCRUALS_{i,t} = \beta_0 + \beta_1 (1/ASSETS_{i,t}) + \beta_2 (\Delta SALES_{i,t} - \Delta REC_{i,t}) + \beta_3 PPE_{i,t} + \varepsilon_{i,t},$$

where for firm i and fiscal year t , $TOTAL\_ACCRUALS$ equals (net income before extraordinary items minus operating cash flows from continuing operations _{t})/average total assets _{$t-1$} ; $ASSETS$ equals average total assets _{$t-1$} ; ΔREC equals (change in accounts receivable from year _{$t-1$} to year _{t})/average total assets _{$t-1$} ; PPE equals net property, plant, and equipment in year/average total assets _{$t-1$} ; and ε equals the pre-K LW estimated discretionary accrual. From each firm's estimated pre-K LW discretionary accrual, we subtract the estimated pre-K LW discretionary accrual of the closest ROA firm in the same two-digit SIC code. The resulting normalized error term is the K LW discretionary accrual measure.

TABLE 1
Discretionary Accruals Analysis: Descriptive Statistics
Full and Propensity-Score Matched Samples

	Full Sample				Propensity-Score Matched Sample: Matched Using the Full Model		
	All Obs. Mean Std. Dev.	Big 4 Mean Std. Dev.	Non-Big 4 Mean Std. Dev.	Difference in Means (t-statistic)	Big 4 Mean Std. Dev.	Non-Big 4 Mean Std. Dev.	Difference in Means (t-statistic)
<i>ADA</i>	0.1031 0.1398	0.0941 0.1193	0.1430 0.2033	-0.0488*** (-36.76)	0.1278 0.1582	0.1312 0.1918	-0.0034 (-1.42)
<i>Total Assets</i>	1,698 5,333	2,021 5,789	258 1,794	1,763*** (34.71)	333 2,042	321 2,006	11.77 (0.42)
<i>Total MKT Value</i>	1,872 7,334	2,239 8,028	234 1,694	2,004*** (28.63)	347 2,158	289 1,895	57.51** (2.06)
<i>LOG_MKT</i>	4.9784 2.3611	5.4111 2.2334	3.0450 1.9013	2.3660*** (110.00)	3.5247 1.9161	3.3265 1.8949	0.1982*** (7.56)
<i>ROA</i>	-0.0273 0.2557	-0.0106 0.2206	-0.1018 0.3652	0.0912*** (37.50)	-0.0822 0.2990	-0.0843 0.3411	0.0021 (0.49)
<i>LEV</i>	0.2159 0.1979	0.2207 0.1967	0.1948 0.2016	0.0261*** (13.62)	0.1994 0.2074	0.1983 0.1988	0.0011 (0.41)
<i>CURR</i>	3.1482 5.5247	3.0324 5.0077	3.6661 7.3841	-0.6336*** (-11.96)	3.5748 5.5612	3.6350 7.6348	-0.0602 (-0.66)
No. Obs.	72,600	59,323	13,277		10,565	10,565	
% of Total	100%	77.6%	22.4%				

*, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels, respectively, using two-tailed t-tests of differences in means.

This table presents the descriptive statistics for our full and propensity-score matched discretionary accruals samples. Propensity scores were calculated using Equation (1).

Variable Definitions:

ADA = absolute discretionary accruals as per Kothari et al. (2005) in year_{*i*};

Total Assets = total assets at the end of year_{*i*};

Total MKT Value = market value of equity at the end of year_{*i*};

LOG_MKT = natural logarithm of market value of equity at the end of year_{*i*};

ROA = income before extraordinary items_{*t-1*}/average total assets_{*t-1*};

LEV = (long-term debt_{*t-1*} plus debt in current liabilities_{*t-1*})/average total assets_{*t-1*}; and

CURR = current assets_{*t-1*}/current liabilities_{*t-1*}.

propensity-score matched sample, except *LOG_MKT*, are insignificantly different at the 10 percent level between the two client types. *LOG_MKT* is still significantly different between the two client groups because the propensity-score model uses total assets rather than market value to proxy for firm size. When we use market value as the proxy for firm size in the propensity-score model, *LOG_MKT* is insignificantly different between the two client groups.

In Table 2, we confirm the results of Becker et al. (1998), Francis et al. (1999), and Butler et al. (2004) for the full sample, as we find a negative and significant difference in means for *BIG4* of -0.0488 ($t = -36.76$; $p < 0.01$), and a negative and significant *BIG4* coefficient of -0.0179 ($t = -8.48$; $p < 0.01$) in the two Full Sample columns, respectively. All control variable coefficients are significant ($p < 0.01$) and have directional effects consistent with those documented by the previous studies. Regressions in all analyses include untabulated year and industry fixed effects.

TABLE 2
Discretionary Accruals Analysis: Univariate and Multivariate Tests
Full and Propensity-Score Matched Samples

$$ADA_{i,t} = \beta_0 + \beta_1 BIG4_{i,t} + \beta_2 LOG\_MKT_{i,t} + \beta_3 ROA_{i,t-1} + \beta_4 LEV_{i,t-1} + \beta_5 CURR_{i,t-1} + Industry\_FE + Year\_FE + \epsilon_{i,t} \quad (2)$$

Dependent Variable: ADA				
	Predicted Sign	Full Sample		Propensity-Score Matched Sample: Matched Using the Full Model
		Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)	Difference in Means (t-statistic)
Intercept			0.1128*** (13.43)	0.0828*** (6.34)
BIG4	—	−0.0488*** (−36.76)	−0.0179*** (−8.48)	−0.0034 (−1.42)
LOG_MKT	—		−0.0059*** (−18.93)	−0.0054*** (−7.99)
ROA	—		−0.1290*** (−18.89)	−0.1319*** (−15.03)
LEV	—		−0.0497*** (−17.08)	−0.0570*** (−9.42)
CURR	—		−0.0003*** (−2.60)	−0.0006*** (−3.43)
Industry_FE			Included	Included
Year_FE			Included	Included
Adjusted R ²			0.14	0.11
No. Obs.		72,600	72,600	21,130
Matching Model R ²				0.29
% Correctly Classified				86.03%

*, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels, respectively, using two-tailed tests.

This table presents the results of our discretionary accruals tests using the full and propensity-score matched samples.

Multivariate estimates are based on Equation (2). Propensity scores were calculated using Equation (1). t-statistics and p-values are calculated using clustered standard errors by firm for the multivariate analyses. For brevity, the year-specific and industry-specific intercepts are not reported. The matching model R² is the pseudo R² for the propensity-score logistic regression. The percentage correctly classified refers to the percentage of audit clients that are correctly classified as Big 4 or non-Big 4 clients, based on a 50 percent cutoff level, using the predicted probabilities from the propensity-score model.

Variable Definitions:

- ADA = absolute discretionary accruals as per Kothari et al. (2005) in year;
- BIG4 = 1 if the client has a Big 4 auditor in year, and 0 otherwise;
- LOG_MKT = natural logarithm of market value of equity at the end of year;
- ROA = income before extraordinary items_{t-1} / average total assets_{t-1};
- LEV = (long-term debt_{t-1} plus debt in current liabilities_{t-1}) / average total assets_{t-1}; and
- CURR = current assets_{t-1} / current liabilities_{t-1}.

The last two columns of Table 2 present the results of the propensity-score matched sample. We find an insignificant difference in means for *BIG4* of -0.0034 ($t = -1.42$; $p = 0.15$) and an insignificant multivariate *BIG4* coefficient of -0.0018 ($t = -0.73$; $p = 0.46$), suggesting that once client characteristics are balanced between the two clienteles, the treatment effects of Big 4 auditors are insignificantly different from those of non-Big 4 auditors with respect to discretionary accruals. All control variable coefficients are significant ($p < 0.01$) and have directional effects consistent with those documented in previous studies.¹²

To examine whether the Big 4 effect can be attributed to specific client characteristics, we perform the following analysis. Using the RESET test, we find that total assets (*LOG_ASSETS*), return on assets (*ROA_{t-1}*), total leverage (*LEV_{t-1}*), and the current ratio (*CURR_{t-1}*) are nonlinear to both the auditor choice and discretionary accruals; hence, we create single-variable matched samples, with year and industry indicators, using each of these foregoing variables. Table 3 reports the discretionary accruals regressions for these matched samples, where columns one and two report results for the *LOG_ASSETS* matched sample, columns three and four report the results for the *ROA_{t-1}* matched sample, columns five and six report the results for the *LEV_{t-1}* matched sample, and columns seven and eight report the results for the *CURR_{t-1}* matched sample. The *LOG_ASSETS* matched sample has statistically insignificant *BIG4* coefficients of -0.0030 ($t = -1.21$; $p = 0.23$) and -0.0026 ($t = -1.03$; $p = 0.30$) for the difference in means and multivariate results, respectively, consistent with the results using the full propensity-score matched sample. However, the coefficients on *BIG4* are negative and significant ($p < 0.01$) for all other single-variable matched samples.

It is not surprising that only the *LOG_ASSETS* matched sample yields the same inferences as the full propensity-score matched sample, given that client size is the characteristic most closely related to auditor selection. For example, in Table 2, the full propensity-score matching model explains 29 percent of the variation in auditor choice (see the matching model R^2); whereas in Table 3, the *LOG_ASSETS* matching model (predicting auditor choice with only total assets, year, and industry) explains 28 percent of the variation in auditor choice. The other single-variable matched models at best only explain 6 percent of the variation in auditor choice. Taken together, the findings from our matched samples suggest that the Big 4 effect, using discretionary accruals as an audit-quality proxy, appears to be attributable to client characteristics, primarily client size.

VI. ANALYSIS 2: EX ANTE COST-OF-EQUITY CAPITAL

Method

Following Khurana and Raman (2004; hereafter KR), we examine the relation between auditor type and the *ex ante* cost of capital, as follows:

$$RPEG_{i,t} = \beta_0 + \beta_1 BIG4_{i,t} + \beta_2 BETA_{i,t} + \beta_3 LOG\_LEV_{i,t} + \beta_4 VAR_{i,t} + \beta_5 LOG\_MKT_{i,t} + \beta_6 LOG\_BTM_{i,t} + \beta_7 GRWTH_{i,t} + Industry\_FE + Year\_FE + \varepsilon_{i,t}, \quad (3)$$

¹² Hribar and Nichols (2007) document that measurement error in absolute discretionary accruals is often correlated with the partitioning variable in tests of earnings management. They specifically identify that the *BIG4* partitioning variable is highly correlated with the following two main drivers of measurement error in absolute discretionary accruals: firm size and operating volatility (Hribar and Nichols 2007, 1033). Following their guidance to correct for measurement error in estimating the partitioning variable, we simultaneously include several firm-size (the logarithm of assets, sales, and market value) and operating-volatility (cash-flow and sales volatility) controls in our discretionary accruals regressions (untabulated), and find similar results in the full and matched samples to those documented in Tables 2 and 3. Moreover, we ran our analyses separately for positive and negative discretionary accruals and again, all inferences are the same as those documented in our main analyses.

TABLE 3
Discretionary Accruals Analysis: Univariate and Multivariate Tests
Attribute-Based Matched Samples

$$ADA_{i,t} = \beta_0 + \beta_1 BIG4_{i,t} + \beta_2 LOG\_MKT_{i,t} + \beta_3 ROA_{i,t-1} + \beta_4 LEV_{i,t-1} + \beta_5 CURR_{i,t-1} + Industry\_FE + Year\_FE + \varepsilon_{i,t} \quad (2)$$

Dependent Variable: ADA

	Size Matched Sample: Matched Using LOG_ASSETS, Year, and Industry		ROA _{i,t-1} Matched Sample: Matched Using ROA _{i,t-1} , Year, and Industry		LEV _{i,t-1} Matched Sample: Matched Using LEV _{i,t-1} , Year, and Industry		CURR _{i,t-1} Matched Sample: Matched Using CURR _{i,t-1} , Year, and Industry	
	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)
BIG4	-0.0030 (-1.21)	-0.0026 (-1.03) 0.11	-0.0321*** (-14.49)	-0.0162*** (-5.52) 0.13	-0.0436*** (-21.16)	-0.0164*** (-6.16) 0.13	-0.0454*** (-22.44)	-0.0174*** (-6.99) 0.13
Adjusted R ²								
No. Obs.	22,210	22,210	26,456	26,456	26,540	26,540	26,552	26,552
Matching Model R ²	0.28		0.06		0.04		0.04	
% Correctly Classified	85.78%		81.89%		81.72%		81.71%	

*, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels, respectively, using two-tailed tests.

This table presents the results of our discretionary accruals tests using four different attribute-based matched samples. We use the propensity-score methodology to implement the attribute-based matching. Propensity scores were calculated using only a single variable with year and industry indicators. Multivariate estimates are based on Equation (2). t-statistics and p-values are calculated using clustered standard errors by firm for the multivariate analyses. For brevity, we only report the coefficients on BIG4. The matching model R² is the pseudo R² for the propensity-score logistic regression. The percentage correctly classified refers to the percentage of audit clients that are correctly classified as Big 4 or non-Big 4 clients, based on a 50 percent cutoff level, using the predicted probabilities from the propensity-score model.

Variable Definitions:

ADA = absolute discretionary accruals as per Kothari et al. (2005) in year;

BIG4 = 1 if the client has a Big 4 auditor in year, and 0 otherwise;

LOG_MKT = natural logarithm of market value of equity at the end of year;

ROA = income before extraordinary items_{i,t-1} / average total assets_{i,t-1};

LEV = (long-term debt_{i,t-1} plus debt in current liabilities_{i,t-1}) / average total assets_{i,t-1}; and

CURR = current assets_{i,t-1} / current liabilities_{i,t-1}.

where for firm i and fiscal year t :

$RPEG$ = *ex ante* cost-of-equity capital estimated using the r_{peg} approach as per Easton (2004) and used by KR;

$BIG4$ = 1 if the client has a Big 4 auditor in year $_t$, and 0 otherwise;

$BETA$ = stock beta (systematic risk) calculated over the 36 months ending in the month of the fiscal year-end $_t$;

LOG_LEV = natural logarithm of (long-term debt $_t$ plus debt in current liabilities $_t$)/average total assets $_{t-1}$;

VAR = earnings variability measured by the dispersion in analysts' earnings forecasts available on I/B/E/S during the fiscal year-end month;

LOG_MKT = natural logarithm of market value of equity at the end of year $_t$;

LOG_BTM = natural logarithm of book value of equity/market value of equity at the end of year $_t$; and

$GRWTH$ = forecasted growth measured as the difference between the mean analysts' two-year- and one-year-ahead earnings forecasts scaled by the one-year-ahead earnings forecast.

For comparative purposes, all variables and their predicted directions are as specified by KR.

Results

Table 4 presents descriptive statistics for both the full and propensity-score matched *ex ante* cost-of-equity capital, $RPEG$, samples. Due to restrictions concerning analyst forecast data, there are 25,068 firm-year observations in the full sample, of which 23,856 (94.9 percent) and 1,212 (5.1 percent) represent Big 4 and non-Big 4 clients, respectively. On average, firms in this full sample are significantly larger across both auditor types than in the accruals full sample: the mean market value and total assets of Big 4 clients are \$4.5 billion and \$4.2 billion, respectively, and the mean market value and total assets of non-Big 4 clients are \$1.5 billion and \$1.9 billion, respectively, indicating that the size differential is smaller in this sample than in the accruals sample. Also worth noting in the $RPEG$ full sample is that Big 4 clients, on average, have significantly lower $RPEG$, $BETA$, and LOG_BTM values than non-Big 4 clients. The propensity-score model once again appears effective in forming a balanced sample of Big 4 and non-Big 4 clients as all the $RPEG$ control variables in the propensity-score matched sample are insignificantly different at the 10 percent level between the two client types.

In Table 5, we first confirm the results of KR for the full sample, as we find a negative and significant difference in means for $BIG4$ of -0.0072 ($t = -4.99$; $p < 0.01$), and a negative and significant $BIG4$ coefficient of -0.0037 ($t = -2.14$, $p = 0.03$) in the two Full Sample columns, respectively. All control variable coefficients are significant ($p < 0.01$) and have directional effects consistent with those documented by KR for their U.S. sample. The last two columns of Table 5 present the results of the propensity-score matched sample. We find an insignificant difference in means for $BIG4$ of -0.0031 ($t = -1.43$; $p = 0.15$) and an insignificant $BIG4$ coefficient of -0.0021 ($t = -1.01$; $p = 0.31$). All the control variable coefficients remain significant and have directional effects consistent with those reported by KR.

Again, to investigate whether the Big 4 effect could reflect client characteristics, we perform the following procedures. Using the RESET test, we find that total assets (LOG_ASSETS), the stock beta ($BETA$), total leverage (LOG_LEV), and the book-to-market ratio (LOG_BTM) are nonlinear to both the auditor choice and to the *ex ante* cost-of-equity capital. Hence, we create single-variable matched samples, with year and industry indicators, using each of these variables. Table 6 reports the $RPEG$ regressions using these attribute-based matched samples, where col-

TABLE 4
Ex Ante Cost-of-Equity Capital Analysis: Descriptive Statistics
Full and Propensity-Score Matched Samples

	Full Sample				Propensity-Score Matched Sample: Matched Using the Full Model		
	All Obs. Mean Std. Dev.	Big 4 Mean Std. Dev.	Non-Big 4 Mean Std. Dev.	Difference in Means (t-statistic)	Big 4 Mean Std. Dev.	Non-Big 4 Mean Std. Dev.	Difference in Means (t-statistic)
<i>RPEG</i>	0.1146 0.0492	0.1142 0.0489	0.1215 0.0537	-0.0072*** (-4.99)	0.1182 0.0530	0.1214 0.0536	-0.0031 (-1.43)
<i>Total Assets</i>	4,093 12,122	4,202 12,258	1,944 8,784	2,258*** (6.33)	1,723 7,802	1,955 8,812	-231.51 (-0.68)
<i>Total MKT Value</i>	4,352 12,342	4,499 12,575	1,466 5,428	3,033*** (8.36)	1,516 5,906	1,474 5,444	42.17 (0.18)
<i>LOG_MKT</i>	6.7469 1.7851	6.7938 1.7837	5.8235 1.5471	0.9702*** (18.59)	5.8871 1.5574	5.8257 1.5518	0.0614 (0.97)
<i>BETA</i>	1.1029 0.7566	1.0959 0.7461	1.2403 0.9292	-0.1443*** (-6.48)	1.1983 0.8072	1.2441 0.9308	-0.0461 (-1.28)
<i>LOG_LEV</i>	-1.8453 1.3257	-1.8283 1.3060	-2.1791 1.6310	0.3510*** (9.01)	-2.2303 1.7432	-2.1860 1.6341	-0.0442 (-0.64)
<i>VAR</i>	0.0687 0.1033	0.0689 0.1037	0.0658 0.0961	0.0030 (1.01)	0.0640 0.0101	0.0658 0.0960	-0.0018 (-0.45)
<i>LOG_BTM</i>	-0.8294 0.6862	-0.8331 0.6883	-0.7565 0.6374	-0.0766*** (-3.79)	-0.7405 0.6928	-0.7591 0.6383	0.0186 (0.69)
<i>GRWTH</i>	0.3314 1.5853	0.3306 1.6051	0.3476 1.1257	-0.0171 (-0.37)	0.4150 2.0604	0.3479 1.1289	0.0671 (0.99)
No. Obs.	25,068	23,856	1,212		1,204	1,204	
% of Total	100%	94.9%	5.1%				

*, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels, respectively, using two-tailed t-tests of differences in means.

This table presents the descriptive statistics for our full and propensity-score matched *ex ante* cost-of-equity capital samples. Propensity scores were calculated using Equation (1).

Variable Definitions:

RPEG = *ex ante* cost-of-equity capital estimated using the r_{peg} approach as per Easton (2004) and used by Khurana and Raman (2004);

Total Assets = total assets at the end of year;

Total MKT Value = market value of equity at the end of year;

LOG_MKT = natural logarithm of market value of equity at the end of year;

BETA = stock beta (systematic risk) calculated over the 36 months ending in the month of the fiscal year-end;

LOG_LEV = natural logarithm of (long-term debt, plus debt in current liabilities,)/average total assets_{*t-1*};

VAR = earnings variability measured by the dispersion in analysts' earnings forecasts available on I/B/E/S during the fiscal year-end month;

LOG_BTM = natural logarithm of book value of equity/market value of equity at the end of year; and

GRWTH = forecasted growth measured as the difference between the mean analysts' two-year- and one-year-ahead earnings forecasts scaled by the one-year-ahead earnings forecast.

TABLE 5
Ex Ante Cost-of-Equity Capital Analysis: Univariate and Multivariate Tests
Full and Propensity-Score Matched Samples

$$RPEG_{i,t} = \beta_0 + \beta_1 BIG4_{i,t} + \beta_2 BETA_{i,t} + \beta_3 LOG\_LEV_{i,t} + \beta_4 VAR_{i,t} + \beta_5 LOG\_MKT_{i,t} \\ + \beta_6 LOG\_BTM_{i,t} + \beta_7 GRWTH_{i,t} + Industry\_FE + Year\_FE + \varepsilon_{i,t} \quad (3)$$

		Dependent Variable: <i>RPEG</i>			
		Full Sample		Propensity-Score Matched Sample: Matched Using the Full Model	
	Predicted Sign	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)
Intercept			0.1558*** (22.09)		0.1517*** (6.15)
<i>BIG4</i>	—	−0.0072*** (−4.99)	−0.0037** (−2.14)	−0.0031 (−1.43)	−0.0021 (−1.01)
<i>BETA</i>	+		0.0099*** (16.51)		0.0091*** (5.29)
<i>LOG_LEV</i>	+		0.0029*** (10.52)		0.0028*** (4.40)
<i>VAR</i>	+		0.0232*** (5.38)		0.0313** (2.14)
<i>LOG_MKT</i>	—		−0.0062*** (−19.77)		−0.0093*** (−8.96)
<i>LOG_BTM</i>	+		0.0105*** (13.71)		0.0048** (1.99)
<i>GRWTH</i>	+		0.0067*** (4.63)		0.0079*** (5.04)
<i>Industry_FE</i>			Included		Included
<i>Year_FE</i>			Included		Included
Adjusted R ²			0.28		0.33
No. Obs.		25,068	25,068	2,408	2,408
Matching Model R ²				0.11	
% Correctly Classified				95.00%	

*, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels, respectively, using two-tailed tests.

This table presents the results of our *ex ante* cost-of-equity capital tests using the full and propensity-score matched samples. Multivariate estimates are based on Equation (3). Propensity scores were calculated using Equation (1). t-statistics and p-values are calculated using clustered standard errors by firm for the multivariate analyses. For brevity, the year-specific and industry-specific intercepts are not reported. The matching model R² is the pseudo R² for the propensity-score logistic regression. The percentage correctly classified refers to the percentage of audit clients that are correctly classified as Big 4 or non-Big 4 clients, based on a 50 percent cutoff level, using the predicted probabilities from the propensity-score model.

Variable Definition:

RPEG = *ex ante* cost-of-equity capital estimated using the r_{peg} approach as per Easlon (2004) and used by Khurana and Raman (2004);

BIG4 = 1 if the client has a Big 4 auditor in year, and 0 otherwise;

(continued on next page)

TABLE 5 (continued)

BETA = stock beta (systematic risk) calculated over the 36 months ending in the month of the fiscal year-end;
LOG_LEV = natural logarithm of (long-term debt, plus debt in current liabilities)/average total assets_{*t-1*};
VAR = earnings variability measured by the dispersion in analysts' earnings forecasts available on I/B/E/S during the fiscal year-end month;
LOG_MKT = natural logarithm of market value of equity at the end of year;_{*t*}
LOG_BTM = natural logarithm of book value of equity/market value of equity at the end of year;_{*t*} and
GRWTH = forecasted growth measured as the difference between the mean analysts' two-year- and one-year-ahead earnings forecasts scaled by the one-year-ahead earnings forecast.

umns one and two report results for the *LOG_ASSETS* matched sample, columns three and four report the results for the *BETA* matched sample, columns five and six report the results for the *LOG_LEV* matched sample, and columns seven and eight report the results for the *LOG_BTM* matched sample.

As in the accruals analysis, the difference in means and the coefficient on *BIG4* for the multivariate results in the *LOG_ASSETS* matched sample are insignificant: -0.0032 ($t = -1.49$; $p = 0.14$) and -0.0013 ($t = -0.61$; $p > 0.50$), respectively. The differences in means for *BIG4* are negative and significant, and the coefficients on *BIG4* in the multivariate results are insignificant for all other single-variable matched samples. As in the discretionary accruals analysis, the *LOG_ASSETS* matched sample yields the same inferences as the full propensity-score matched sample. Together, these findings suggest that the Big 4 effect on the *ex ante* cost-of-equity capital are likely attributable to client characteristics, primarily client size.

VII. ANALYSIS 3: ANALYST FORECAST ACCURACY

Method

To test whether the Big 4 versus non-Big 4 differences in analyst forecast accuracy can be attributed to client characteristics, we use the following model employed by Behn et al. (2008; hereafter BCK):

$$\begin{aligned} ACCY_{i,t} = & \beta_0 + \beta_1 BIG4_{i,t} + \beta_2 LOG\_MKT_{i,t} + \beta_3 SURPRISE_{i,t} + \beta_4 NETLOSS_{i,t} + \beta_5 ZMIJ_{i,t} \\ & + \beta_6 HORIZON_{i,t} + \beta_7 STDROE_{i,t} + \beta_8 NANA_{i,t} + \beta_9 EL_{i,t} + Industry\_FE + Year\_FE \\ & + \varepsilon_{i,t}, \end{aligned} \quad (4)$$

where for firm *i* and fiscal year *t*:

ACCY = negative of the absolute value of the difference between analysts' earnings forecasts of eps_{t+1} and actual eps_{t+1} , scaled by stock price at the end of year_{*t*}, as per Lang and Lundholm (1996) and used by BCK;

BIG4 = 1 if the client has a Big 4 auditor in year_{*t*}, and 0 otherwise;

LOG_MKT = natural logarithm of market value of equity at the end of year_{*t*};

SURPRISE = (net income_{*t+1*} – net income_{*t*})/market value of equity at the end of year_{*t*};

NETLOSS = 1 if client has negative net income_{*t+1*}, and 0 otherwise;

ZMIJ = distress score calculated using Zmijewski's (1984) unweighted original parameters;

HORIZON = natural logarithm of the average number of calendar days between forecast announcement date and subsequent earnings announcement date;

STDROE = standard deviation of net income over the five years from year_{*t-4*} until year_{*t*};

NANA = natural logarithm of the number of analysts following the client; and

TABLE 6
Ex Ante Cost-of-Equity Capital Analysis: Univariate and Multivariate Tests
 Attribute-Based Matched Samples

$$RPEG_{i,t} = \beta_0 + \beta_1 BIG4_{i,t} + \beta_2 BETA_{i,t} + \beta_3 LOG\_LEV_{i,t} + \beta_4 VAR_{i,t} + \beta_5 LOG\_MKT_{i,t} + \beta_6 LOG\_BTM_{i,t} + \beta_7 GRWTH_{i,t} + Industry\_FE + Year\_FE + \varepsilon_{i,t} \quad (3)$$

	Dependent Variable: RPEG					
	Size Matched Sample: Matched Using LOG_ASSETS, Year and Industry		BETA Matched Sample: Matched Using BETA, Year and Industry		LOG LEV Matched Sample: Matched Using LOG_LEV, Year and Industry	
	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)
BIG4	-0.0032 (-1.49)	-0.0013 (-0.61)	-0.0087*** (-4.18)	-0.0011 (-0.54)	-0.0119*** (-5.96)	-0.0025 (-1.32)
Adjusted R ²		0.33		0.33		0.32
No. Obs.	2,408	2,408	2,408	2,408	2,408	2,408
Matching Model R ²	0.11		0.05		0.06	
% Correctly Classified	95.02%		95.01%		95.01%	
					0.05	2,408
					95.01%	95.01%

*, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels, respectively, using two-tailed tests.

This table presents the results of our *ex ante* cost-of-equity capital tests using four different attribute-based matched samples. We use the propensity-score methodology to implement the attribute-based matching. Propensity scores were calculated using only a single variable with year and industry indicators.

Multivariate estimates are based on Equation (3). t-statistics and p-values are calculated using clustered standard errors by firm for the multivariate analyses.

For brevity, we only report the coefficients on BIG4. The matching model R² is the pseudo R² for the propensity-score logistic regression. The percentage correctly classified refers to the percentage of audit clients that are correctly classified as Big 4 or non-Big 4 clients, based on a 50 percent cutoff level, using the predicted probabilities from the propensity-score model.

Variable Definitions:

RPEG = *ex ante* cost-of-equity capital estimated using the r_{peg} approach as per Easton (2004) and used by Khurana and Raman (2004);

BIG4 = 1 if the client has a Big 4 auditor in year, and 0 otherwise;

(continued on next page)

TABLE 6 (continued)

BETA = stock beta (systematic risk) calculated over the 36 months ending in the month of the fiscal year-end;
LOG_LEV = natural logarithm of (long-term debt, plus debt in current liabilities)/average total assets;_{*t-1*}
VAR = earnings variability measured by the dispersion in analysts' earnings forecasts available on I/B/E/S during the fiscal year-end month;
LOG_MKT = natural logarithm of market value of equity at the end of year;_{*t*}
LOG_BTMT = natural logarithm of book value of equity at the end of year;_{*t*} and
GRWTH = forecasted growth measured as the difference between the mean analysts' two-year- and one-year-ahead earnings forecasts scaled by the one-year-ahead earnings forecast.



$$EL = \text{actual } \text{eps}_{t+1}.$$

For comparative purposes, all variables and their predicted directions are as specified by BCK.

Results

Table 7 presents descriptive statistics for both the full and propensity-score matched analyst forecast accuracy, *ACCY*, samples. The full sample has a total of 28,037 firm-year observations, of which 26,521 (94.3 percent) and 1,516 (5.7 percent) represent Big 4 and non-Big 4 clients, respectively. As with the *ex ante* cost-of-equity capital analysis, the firms in this full sample are significantly larger across both client types than in the accruals full sample and, hence, the size differential between the two groups is again smaller in this sample than it is in the accruals analysis. Overall, the descriptive statistics again confirm that Big 4 and non-Big 4 clienteles are significantly different in terms of size and profitability. Moreover, relevant to this analysis, we find that Big 4 audit clients have a larger analyst following and a lower probability of bankruptcy than non-Big 4 audit clients.

In Table 8, we first confirm the results of BCK for the full sample, as we find a positive and significant difference in means for *BIG4* of 0.0083 ($t = 7.21$; $p < 0.01$), and a positive and significant *BIG4* coefficient of 0.0042 ($t = 2.34$; $p = 0.02$) in the two Full Sample columns, respectively. Except for *HORIZON*, *NANA*, and *EL*, all control variable coefficients have directional effects and significance levels consistent with those documented by BCK. The last two columns of Table 8 present the results of the propensity-score matched sample. We find an insignificant difference in means for *BIG4* of 0.0026 ($t = 1.30$; $p = 0.19$) and an insignificant *BIG4* coefficient of 0.0031 ($t = 1.54$; $p = 0.12$).

Using the RESET test, we find that the total assets (*LOG_ASSETS*), the standard deviation of the return on equity (*STDROE*), and analyst following (*NANA*) are nonlinear to both the auditor choice and to analyst forecast accuracy. Hence, we create single-variable matched samples, with year and industry indicators, using each of these variables. Table 9 reports the *ACCY* regressions for the attribute-based matched samples, where columns one and two report results for the *LOG_ASSETS* matched sample, columns three and four report the results for the *STDROE* matched sample, and columns five and six report the results for the *NANA* matched sample.

As in the accruals and *ex ante* cost-of-equity capital analyses, the difference in means and the coefficients on *BIG4* for the multivariate results in the *LOG_ASSETS* matched sample are insignificant: -0.0021 ($t = -0.91$; $p = 0.36$) and 0.0013 ($t = 0.59$; $p > 0.50$), respectively. For the *STDROE* matched sample, the difference in means for *BIG4* is positive and significant, while the *BIG4* coefficient in the multivariate analyses is insignificant. Moreover, the difference in means and the coefficients on *BIG4* in the *NANA* matched sample are positive and significant. In line with the other two analyses, these findings suggest that client characteristics, and particularly client size, could potentially confound the inferences pertaining to the Big 4 effect.

Our findings concerning analyst forecast accuracy appear to be consistent with a 2008 CFA Institute Center survey of 617 CFA investment analysts documenting that the majority of analysts do not prefer Big 4 auditors to non-Big 4 auditors. Specifically, only 41 percent of the respondents generally indicated that they had a preference for firms using “brand-name” auditors; moreover, only 15 percent of the respondents thought that the attractiveness of a company as an investment is detracted when a smaller company switches to a lower-cost auditor that may be more efficient and cost-effective (CFA Institute Center 2008).

VIII. SENSITIVITY ANALYSES

Bootstrapping, Kernel Weighting, and Random Subsamples

To mitigate concerns that our findings are a consequence of the smaller sample sizes, we

TABLE 7
Analyst Forecast Accuracy Analysis: Descriptive Statistics
Full and Propensity-Score Matched Samples

	Full Sample				Propensity-Score Matched Sample: Matched Using the Full Model		
	All Obs. Mean Std. Dev.	Big 4 Mean Std. Dev.	Non-Big 4 Mean Std. Dev.	Difference in Means (t-statistic)	Big 4 Mean Std. Dev.	Non-Big 4 Mean Std. Dev.	Difference in Means (t-statistic)
<i>ACCY</i>	-0.0122 0.0437	-0.0117 0.0425	-0.0201 0.0608	0.0083*** (7.21)	-0.0162 0.0511	-0.0189 0.0593	0.0027 (1.30)
<i>Total Assets</i>	4,243 12,787	4,377 12,990	1,893 8,141	2,484*** (7.36)	1,382 5,527	1,929 8,245	-547** (-2.12)
<i>Total MKT Value</i>	4,875 17,331	5,058 17,729	1,677 6,756	3,381*** (7.39)	1,238 4,367	1,708 6,839	-470** (-2.22)
<i>LOG_MKT</i>	6.6934 1.8677	6.7582 1.8508	5.5594 1.7975	1.1989*** (24.57)	5.6056 1.6483	5.5848 1.7950	0.0208 (0.33)
<i>SURPRISE</i>	0.0814 0.1917	0.0812 0.1909	0.0838 0.2062	-0.0026 (-0.50)	0.0865 0.2198	0.0821 0.2040	0.0045 (0.58)
<i>NETLOSS</i>	0.2036 0.4027	0.2029 0.4021	0.2170 0.4123	-0.0141* (-1.33)	0.2203 0.4146	0.2108 0.4080	0.0095 (0.63)
<i>ZMIJ</i>	-3.0792 1.3155	-3.0710 1.3085	-3.2154 1.4266	0.1441*** (4.15)	-3.1456 1.4487	-3.2279 1.4024	0.0822 (1.57)
<i>HORIZON</i>	3.3829 0.5051	3.3833 0.5029	3.3765 0.5426	0.0068 (0.51)	3.3892 0.5385	3.3766 0.5372	0.0126 (0.74)
<i>STDROE</i>	0.3623 1.1282	0.3486 1.0816	0.6025 1.7360	-0.2539*** (-8.53)	0.5285 1.4723	0.5580 1.6361	-0.0295 (-0.52)
<i>NANA</i>	0.9265 0.8588	0.9457 0.8626	0.5905 0.7106	0.3552*** (15.73)	0.5940 0.6943	0.5985 0.7137	-0.0045 (-0.17)
<i>EL</i>	0.9682 2.0088	0.9798 2.0342	0.7661 1.4804	0.2137*** (4.03)	0.7489 1.5335	0.7862 1.4806	-0.0372 (-0.66)
No. Obs.	28,037	26,521	1,516		1,475	1,475	
% of Total	100%	94.3%	5.7%				

*, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels, respectively, using two-tailed t-tests of differences in means.

This table presents the descriptive statistics for our full and propensity-score matched analyst forecast-accuracy samples. Propensity scores were calculated using Equation (1).

Variable Definitions:

ACCY = negative of the absolute value of the difference between analysts' earnings forecasts of eps_{t+1} and actual eps_{t+1} , scaled by stock price at the end of year_{*t*}, as per Lang and Lundholm (1996) and used by Behn et al. (2008);

Total Assets = total assets at the end of year_{*t+1*};

Total MKT Value = market value of equity at the end of year_{*t+1*};

LOG_MKT = natural logarithm of market value of equity at the end of year_{*t+1*};

SURPRISE = (net income_{*t+1*} - net income_{*t*})/market value of equity at the end of year_{*t*};

NETLOSS = 1 if the client has negative net income_{*t+1*}, and 0 otherwise;

ZMIJ = distress score calculated using Zmijewski's (1984) unweighted original parameters;

HORIZON = natural logarithm of the average number of calendar days between the forecast announcement date and the subsequent earnings announcement date;

STDROE = standard deviation of net income over the five years from year_{*t-4*} until year_{*t*};

EL = actual eps_{t+1} ; and

NANA = natural logarithm of the number of analysts following the client.

TABLE 8
Analyst Forecast Accuracy Analysis: Univariate and Multivariate Tests
Full and Propensity-Score Matched Samples

$$\begin{aligned}
 ACCY_{i,t} = & \beta_0 + \beta_1 BIG4_{i,t} + \beta_2 LOG\_MKT_{i,t} + \beta_3 SURPRISE_{i,t} + \beta_4 NETLOSS_{i,t} + \beta_5 ZMIJ_{i,t} \\
 & + \beta_6 HORIZON_{i,t} + \beta_7 STDROE_{i,t} + \beta_8 NANA_{i,t} + \beta_9 EL_{i,t} \\
 & + Industry\_FE + Year\_FE + \varepsilon_{i,t}
 \end{aligned}
 \tag{4}$$

		Dependent Variable: <i>ACCY</i>			
		Full Sample		Propensity-Score Matched Sample: Matched Using the Full Model	
	Predicted Sign	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)
Intercept			−0.0414*** (−9.08)		−0.0302*** (−2.59)
<i>BIG4</i>	+	0.0083*** (7.21)	0.0042** (2.34)	0.0026 (1.30)	0.0031 (1.54)
<i>LOG_MKT</i>	+		0.0027*** (10.12)		0.0052*** (4.36)
<i>SURPRISE</i>	−		−0.0598*** (−13.43)		−0.0439*** (−4.46)
<i>NETLOSS</i>	−		−0.0072*** (−6.87)		−0.0198*** (−4.68)
<i>ZMIJ</i>	−		−0.0028*** (−6.69)		−0.0024* (−1.79)
<i>HORIZON</i>	−		0.0001 (0.22)		−0.0003 (−0.17)
<i>STDROE</i>	−		−0.0021*** (−4.07)		−0.0035*** (−2.66)
<i>NANA</i>	+		0.0016*** (4.25)		−0.0006 (−0.27)
<i>EL</i>	?		0.0006** (2.38)		−0.0006 (−0.62)
<i>Industry_FE</i>			Included		Included
<i>Year_FE</i>			Included		Included
Adjusted R ²			0.19		0.18
No. Obs.		28,037	28,037	2,950	2,950
Matching Model R ²				0.14	
% Correctly Classified				94.62%	

*, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels, respectively, using two-tailed tests.

This table presents the results of our analyst forecast accuracy tests using the full and propensity-score matched samples. Multivariate estimates are based on Equation (4). Propensity scores were calculated using Equation (1).

Following Behn et al. (2008), t-statistics and p-values are calculated using White's (1980) heteroscedasticity-corrected standard errors clustered by firm for the multivariate analyses. For brevity, the year-specific and industry-specific intercepts

(continued on next page)

TABLE 8 (continued)

are not reported. The matching model R^2 is the pseudo R^2 for the propensity-score logistic regression. The percentage correctly classified refers to the percentage of audit clients that are correctly classified as Big 4 or non-Big 4 clients, based on a 50 percent cutoff level, using the predicted probabilities from the propensity-score model.

Variable Definitions:

- ACCY* = negative of the absolute value of the difference between analysts' earnings forecasts of eps_{t+1} and actual eps_{t+1} , scaled by stock price at the end of year_{*t*}, as per Lang and Lundholm (1996) and used by Behn et al. (2008);
- BIG4* = 1 if the client has a Big 4 auditor in year_{*t*}, and 0 otherwise;
- LOG_MKT* = natural logarithm of market value of equity at the end of year_{*t+1*};
- SURPRISE* = (net income_{*t+1*} – net income_{*t*})/market value of equity at the end of year_{*t*};
- NETLOSS* = 1 if the client has negative net income_{*t+1*}, and 0 otherwise;
- ZMIJ* = distress score calculated using Zmijewski's (1984) unweighted original parameters;
- HORIZON* = natural logarithm of the average number of calendar days between the forecast announcement date and the subsequent earnings announcement date;
- STDROE* = standard deviation of net income over the five years from year_{*t-4*} until year_{*t*};
- EL* = actual eps_{t+1} ; and
- NANA* = natural logarithm of the number of analysts following the client.

perform the following analyses. First, for each matched sample in our main analyses, we obtain bootstrap estimates, using 10,000 repetitions. Second, using each audit-quality proxy, we run analyses for the full sample, putting more weight and using kernel weighting (Heckman et al. 1997), on the observations that are closer in the propensity score or client size. We find that all inferences from these tests are the same as those documented in our matched sample analyses.

Moreover, to ensure that the Big 4 effect can still be detected in smaller subsamples, we perform the following analyses. First, for each audit-quality proxy, we obtain bootstrap estimates using 10,000 repetitions and sample sizes equal to those of the matched sample in each repetition. Second, for each audit-quality proxy, we initially select a random subsample with the same number of observations and composition (i.e., 50 percent Big 4 clients and 50 percent Non-Big 4 clients) as in the matched sample. Using this random subsample, we obtain bootstrap estimates using 10,000 repetitions.¹³ We find that the Big 4 effect holds in these alternative analyses.

Varying the Proportion of Big 4 and Non-Big 4 Clients in the Full and Matched Samples

Some of our audit-quality proxies result in samples that have different auditor compositions than the full population, potentially biasing our inferences. Hence, for each audit-quality proxy, we select a random sample with an 80/20 ratio between Big and non-Big 4 auditors, as in the full Compustat population, from both our full and matched samples, and obtain bootstrap estimates using 10,000 repetitions.¹⁴ All reported inferences are robust to these alternative specifications.

Additional Client and Auditor Characteristics

Facing a similar problem of separating the effects of client from those of audit firm characteristics, Francis and Yu (2009; hereafter FY) document that the audit quality of larger Big 4

¹³ For example, the discretionary accruals propensity-score matched sample has 10,565 clients of each auditor type; hence, we randomly select 10,565 of each auditor type from our full sample and estimate our regression Equation (2) for each random sample, repeating this procedure 10,000 times to obtain bootstrapped estimates.

¹⁴ For example, from the discretionary accruals full sample, we randomly draw 40,000 observations from Big 4 clients and 10,000 observations from non-Big 4 clients, achieving an 80/20 ratio. Similarly, from the discretionary accruals matched sample, we randomly draw 8,000 observations from Big 4 clients and 2,000 observations from non-Big 4 clients achieving an 80/20 ratio. Using these two subsamples, one derived from the full sample and the other from the matched sample, we re-estimate our regression models with 10,000 bootstrap repetitions.

TABLE 9
Analyst Forecast Accuracy Analysis: Univariate and Multivariate Tests
Attribute-Based Matched Samples

$$\begin{aligned}
 ACCY_{i,t} = & \beta_0 + \beta_1 BIG4_{i,t} + \beta_2 LOG\_MKT_{i,t} + \beta_3 SURPRISE_{i,t} + \beta_4 NETLOSS_{i,t} \\
 & + \beta_5 ZMIJ_{i,t} + \beta_6 HORIZON_{i,t} + \beta_7 STDROE_{i,t} + \beta_8 NANA_{i,t} + \beta_9 EL_{i,t} \\
 & + Industry\_FE + Year\_FE + \varepsilon_{i,t}
 \end{aligned} \tag{4}$$

	Dependent Variable: <i>ACCY</i>					
	Size Matched Sample: Matched Using <i>LOG_ASSETS</i> , Year and Industry		<i>STDROE</i> Matched Sample: Matched Using <i>STDROE</i> , Year and Industry		<i>NANA</i> Matched Sample: Matched Using <i>NANA</i> , Year and Industry	
	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)	Difference in Means (t-statistic)	Multivariate Estimate (t-statistic)
<i>BIG4</i>	-0.0021 (-0.91)	0.0013 (0.59)	0.0078*** (4.08)	0.0024 (1.19)	0.0105*** (6.13)	0.0038*** (2.10)
Adjusted R ²		0.21		0.22		0.20
No. Obs.	2,956	2,956	2,982	2,982	2,960	2,960
Matching Model R ²	0.13		0.05		0.06	
% Correctly Classified	94.60%		94.55%		94.56%	

*, **, *** Indicate significance at the 0.10, 0.05 and 0.01 levels, respectively, using two-tailed tests.

This table presents the results of our analyst forecast accuracy tests using three different attribute-based matched samples. We use the propensity-score methodology to implement the attribute-based matching. Propensity scores were calculated using only a single variable with year and industry indicators.

Multivariate estimates are based on Equation (4). Following Behn et al. (2008), t-statistics and p-values are calculated using White's (1980) heteroscedasticity-corrected standard errors clustered by firm for the multivariate analyses. For brevity, we only report the coefficients on *BIG4*. The matching model R² is the pseudo R² for the propensity-score logistic regression. The percentage correctly classified refers to the percentage of audit clients that are correctly classified as Big 4 or non-Big 4 clients, based on a 50 percent cutoff level, using the predicted probabilities from the propensity-score model.

Variable Definitions:

ACCY = negative of the absolute value of the difference between analysts' earnings forecasts of eps_{t+1} and actual eps_{t+1} , scaled by stock price at the end of year, as per Lang and Lundholm (1996) and used by Behn et al. (2008);

BIG4 = 1 if the client has a Big 4 auditor in year, and 0 otherwise;

LOG_MKT = natural logarithm of market value of equity at the end of year_{*t+1*};

SURPRISE = (net income_{*t+1*} - net income_{*t*})/market value of equity at the end of year;

NETLOSS = 1 if the client has negative net income_{*t+1*}, and 0 otherwise;

ZMIJ = distress score calculated using Zmijewski's (1984) unweighted original parameters;

HORIZON = natural logarithm of the average number of calendar days between the forecast announcement date and the subsequent earnings announcement date;

STDROE = standard deviation of net income over the five years from year_{*t-4*} until year_{*t*};

EL = actual eps_{t+1} ; and

NANA = natural logarithm of the number of analysts following the client.

offices is superior to that of smaller Big 4 offices. We enhance our study's controls by estimating all our propensity-score and audit-quality regressions including FY's comprehensive list of control variables.¹⁵ We find that the coefficients on *BIG4* in the full and matched samples are insignificant for all three audit-quality proxies. These results (untabulated) provide additional support for the argument that the Big 4 distinction may reflect client and not auditor characteristics.¹⁶

Auditor Office Size and Client Characteristics

FY's tests comparing audit quality among large and small auditor offices may encounter similar difficulties as the Big 4 auditor tests, given that there is a strong positive correlation between office size and client size. To investigate this possibility, we first replicate and confirm FY's study using their full absolute discretionary accruals, small profit increases, and going-concern opinions models, and their sample years. Second, we create an indicator variable, *MED-OFFICE*, equal to 1 for offices above the median office size, and 0 otherwise, in order to match clients of large offices to those of small offices. Third, we confirm FY's findings by replacing *LOGOFFICE* with *MEDOFFICE* in the full samples for the three audit-quality proxies. Finally, we use both client-size and propensity-score matching to match clients of offices above the median office size to those clients of offices below the median office size. We then find an insignificant difference in *MEDOFFICE* between large and small offices for three foregoing audit-quality proxies ($p > 0.50$, $p = 0.29$, and $p = 0.20$, respectively). These matched-sample results are robust to 10,000 bootstrap repetitions. We caution that this analysis does not rule out FY's conclusion that large offices have higher audit quality, although it suggests the importance of fully controlling for client characteristics in tests of audit quality.

Nonlinear Robustness Checks

To some degree, matching models purge the impact of nonlinearities on auditor size in our analyses. Nonetheless, as a robustness check, we use a backfitting partial linear model (Hastie and Tibshirani 1990) to estimate both the linear and nonlinear terms previously identified using the RESET test.¹⁷ We find that all inferences pertaining to the full and matched samples using partial linear regressions are the same as those documented in the main analyses using ordinary least-squares (OLS) regressions.

¹⁵ These variables are as follows: *LOGOFFICE* is the natural logarithm of practice office size based on the aggregated client audit fees (in \$ millions) of a practice office in a specific fiscal year; *INFLUENCE* is the ratio of a specific client's total fees (audit fees plus nonaudit fees) relative to aggregate annual fees generated by the practice office that audits the client; *TENURE* is 1 if auditor tenure is three years or less, and 0 otherwise; *NATIONAL_LEADER* is 1 if an auditor is the number one auditor in an industry in terms of aggregated audit fees in a specific fiscal year, and 0 otherwise; *CITY_LEADER* is 1 if an office is the number one auditor in terms of aggregated client audit fees in an industry within that city in a specific fiscal year, and 0 otherwise; *OPSEG* is the number of operating segments reported in the Compustat segments database for the firm in year; *GEOSEG* is the number of geographic segments reported in the Compustat segments database for the firm in year; *SALESGROWTH* is the one-year growth rate of a firm's sales revenue; *SALESVOLATILITY* is the standard deviation of sales revenue, we use a rolling window and require a minimum of three years of data; *CFOVOLATILITY* is the standard deviation of *CFO*, we use a rolling window and require a minimum of three years of data to estimate; *CFO* is operating cash deflated by lagged total assets; *LOSS* is 1 if net income before extraordinary items is negative, and 0 otherwise; *BANKRUPTCY* is the Altman Z-score, which is a measure of the probability of bankruptcy, with a lower value indicating greater financial distress; *VOLATILITY* is a client's stock volatility and is the standard deviation of 12 monthly stock returns for the current fiscal year; and *MB* is the natural logarithm of the ratio of a client's market value of equity to its book value of equity in year.

¹⁶ When we include the extensive list of control variables as per FY, the Big 4 effect could disappear due to the multicollinearity of these auditor-related variables with the *BIG4* variable. For example, the correlation between *BIG4* and *LOGOFFICE* in our separate analyses is approximately 0.4. Thus, what we learn from the matched sample findings is that they are not likely caused by the multicollinearity of the additional auditor controls with the *BIG4* variable as the matched samples balance and, hence, mitigate these multicollinearity effects to some extent.

¹⁷ A backfitting algorithm simultaneously estimates the effects of nonlinear terms and the coefficients for the linear terms. This algorithm is available in the R and S-Plus packages using the function *gam* (generalized additive model).

In addition to using the RESET test, we generate scatter plots with both linear and local polynomial fits of each nonlinear variable. These scatter plots support the RESET test conclusions concerning the documented nonlinearities in each audit-quality proxy sample. We also run a bootstrap analysis on the RESET analysis drawing 10,000 samples of 5,000 observations, and estimate the RESET tests using the bootstrapped estimates. We find that client size, measured as *LOG_ASSETS* or *LOG_MKT*, is the only variable nonlinear to both the auditor choice and audit-quality proxy in each of the audit-quality analyses. However, there could be other forms of nonlinearity that are not detected by these tests.

Heckman Self-Selection Model

To examine whether the results are sensitive to specifications that consider the Heckman (1979) procedure, we model the self-selection of auditors following Chaney et al. (2004).¹⁸ We estimate the model separately for each of the three audit-quality proxy samples and include the inverse Mills' ratio, estimated by year, in the respective second-stage pooled regressions, in the full and matched samples.¹⁹ In line with Francis et al.'s (2010) critique of the Heckman (1979) model pertaining to satisfying exclusion restrictions, we estimate the inverse Mills ratio controlling for client size using total assets, total sales, market value, and various combinations of these variables. We continue to find evidence of the Big 4 effect in our full samples and do not find an effect for our matched samples using these alternative specifications. Current research is attempting to extend the literature and resolve the differences in estimating treatment effects using either the Heckman (1979) or the propensity-score approaches (Guo and Fraser 2010).

Audit Fees and Client Characteristics

Several studies use audit fees as a proxy to assess the potential superiority of Big 4 auditors, generally finding that the audit fees of the Big 4 clients are significantly larger than those of the non-Big 4 clients.²⁰ Among other factors, fees proxy for the level of service provided (e.g., Davis et al. 1993; Whisenant et al. 2003) and are negatively associated with levels of earnings management (e.g., Frankel et al. 2002; Ashbaugh et al. 2003). However, the existence of a fee premium in itself does not imply higher audit quality, especially if the Big 4 auditors have more pricing power over their clients than do the non-Big 4 auditors (Francis 2004). Moreover, O'Keefe et al. (1994, 242) caution that "inferences about prices in such studies can be erroneous if the cross-sectional variations in auditor effort caused by differences in client characteristics are not adequately controlled."

As a sensitivity analysis, we apply our matching methodology to examine whether the Big 4 premium may be attributed to client characteristics. Using an audit fee model similar to that employed by Chaney et al. (2004), we perform the same set of analyses (untabulated) as with the prior three audit-quality proxies, using Compustat fee data from 2000 to 2006. We find that the

¹⁸ Following Chaney et al. (2004), we model the self-selection of auditors, estimated separately for each year, as follows in order to obtain the inverse Mills ratio:

$$BIG4_i = \beta_0 + \beta_1 LOG\_ASSETS_i + \beta_2 ATURN_i + \beta_3 CURR_i + \beta_4 DTA_i + \beta_5 ROA_i + \beta_6 ROALOSS_i + \varepsilon_i,$$

where *BIG4* equals 1 if the client has a Big 4 Auditor, and 0 otherwise; *LOG_ASSETS* equals the natural logarithm of total assets; *ATURN* equals sales/total assets; *CURR* equals current assets/current liabilities; *DTA* equals total debt/total assets; *ROA* equals net income before extraordinary items/average total assets_{*t-1*}; and *ROALOSS* equals *ROA* multiplied by 1 if client has negative net income, and 0 otherwise.

¹⁹ Other articles in this literature that apply the Heckman model to panel data in this manner include Khurana and Raman (2004), and Behn et al. (2008).

²⁰ Such studies include: Simunic (1980), Francis (1984), Francis and Stokes (1986), Chan et al. (1993), Craswell et al. (1995), DeFond et al. (2000), Ferguson et al. (2003), Whisenant et al. (2003), and Chaney et al. (2004). For an extensive review of this literature, see Hay et al. (2006).

coefficient on *BIG4* is positive and statistically significant at the 5 percent level for the full, the propensity-score matched, and the attribute-based matched samples.²¹ Our results are consistent with the propensity-score matched fee results presented in Clatworthy et al. (2009), and could be a reflection of the following circumstances noted in Simunic (1980) and O'Keefe et al. (1994): (1) that Big 4 auditors provide a higher quantity of audit services; (2) that Big 4 auditors charge a higher unit price; and (3) it is possible that we still do not adequately control for all the relevant client characteristics that drive the quantity of audit services rendered. Our analyses are robust to the inclusion of a comprehensive list of client and auditor characteristics as in Francis and Yu (2009), and to different matching specifications.

IX. CONCLUSION

In this study, we examine whether differences in quality between Big 4 and non-Big 4 audit firms could be a reflection of client characteristics. Using matching models or controlling for an extensive list of client and auditor variables, we find that the treatment effects of Big 4 auditors are insignificantly different from those of non-Big 4 auditors with respect to discretionary accruals, the *ex ante* cost-of-equity capital, and analyst forecast accuracy.

We caution the reader that our findings must be interpreted with due regard to their methodological limitations. First, an inherent limitation of this approach is that we are unable to match on pre-treatment attributes. Second, the propensity-score and attribute-based matching models rely on the assumption that the effects of unobservables are not pertinent to estimating the treatment effects. Third, matching reflects a trade-off between identifying the treatment effects and the ability to generalize results to the full population. Fourth, we cannot ensure that we include all relevant client and auditor control variables. We cannot rule out the possibility that the foregoing limitations could introduce biases into our analyses.

We reemphasize that our study does not resolve the underlying question as to whether differences in audit-quality proxies between Big 4 and non-Big 4 auditors can be attributed to client characteristics, but rather that it only provides suggestive evidence. We hope that our results encourage other researchers to explore alternative methodologies that further disentangle client characteristics from audit-quality effects.

REFERENCES

- Armstrong, C., A. Jagolinzer, and D. Larcker. 2010. Chief executive officer equity incentives and accounting irregularities. *Journal of Accounting Research* 48 (2): 225–271.
- Ashbaugh, H., R. LaFond, and B. W. Mayhew. 2003. Do nonaudit services compromise auditor independence? Further evidence. *The Accounting Review* 78 (3): 611–639.

²¹ Following Chaney et al. (2004), we model the determination of audit fees as follows:

$$\begin{aligned} \text{LOGFEES}_{i,t} = & \beta_0 + \beta_1 \text{BIG4}_{i,t} + \beta_2 \text{LOG\_ASSETS}_{i,t} + \beta_3 \text{ATURN}_{i,t} + \beta_4 \text{EXPORT}_{i,t} + \beta_5 \text{CURR}_{i,t} + \beta_6 \text{LEV}_{i,t} \\ & + \beta_7 \text{QUICK}_{i,t} + \beta_8 \text{ROA}_{i,t} + \beta_9 \text{ROA\_LOSS}_{i,t} + \beta_{10} \text{ABSEXTRA}_{i,t} + \text{Industry\_FE} + \text{Year\_FE} + \varepsilon_{i,t}, \end{aligned}$$

where *LOGFEES* equals the natural logarithm of total audit fees in year_{*t*}; *BIG4* equals 1 if the client has a Big 4 auditor in year_{*t*}, and 0 otherwise; *LOG_ASSETS* equals the natural logarithm of total assets at the end of year_{*t*}; *ATURN* equals sales/total assets_{*t-1*}; *EXPORT* equals the sum of sales from foreign segments in year_{*t*}/total sales in year_{*t*}; *CURR* equals current assets/current liabilities; *LEV* equals (long-term debt, plus debt in current liabilities)/total assets_{*t-1*}; *QUICK* equals (current assets, minus inventory)/current liabilities; *ROA* equals net income_{*t-1*}/total assets_{*t-1*}; *ROA_LOSS* equals *ROA* multiplied by 1 if client has negative net income_{*t-1*}, and 0 otherwise; and *ABSEXTRA* is the absolute value of extraordinary items/total assets_{*t*}. We find that inferences are robust to the inclusion of nonaudit fees and interaction terms between *BIG4*, complexity, and foreign segment proxies.

- Beatty, R. 1989. Auditor reputations and the pricing of initial public offerings. *The Accounting Review* 64 (4): 693–709.
- Becker, C., M. DeFond, J. Jiambalvo, and K. R. Subramanyam. 1998. The effect of audit quality on earnings management. *Contemporary Accounting Research* 15 (1): 1–24.
- Behn, B., J. H. Choi, and T. Kang. 2008. Audit quality and properties of analyst earnings forecasts. *The Accounting Review* 83 (2): 327–359.
- Beneish, M. 1999. Incentives and penalties related to earnings overstatements that violate GAAP. *The Accounting Review* 74 (4): 425–457.
- Butler, M., A. Leone, and M. Willenborg. 2004. An empirical analysis of auditor reporting and its association with abnormal accruals. *Journal of Accounting and Economics* 37 (2): 139–165.
- CFA Institute Center. 2008. *June Survey Where Do You Stand? Results on Auditors*. Charlottesville, VA: CFA Institute.
- . 2009. DG MARKT: Consultation on control structures in audit firms and their consequences on the audit market. Working paper, CFA Institute.
- Chan, P., M. Ezzamel, and D. Gwilliam. 1993. Determinants of audit fees for quoted UK companies. *Journal of Business Finance & Accounting* 20 (6): 765–786.
- Chaney, P., D. Jeter, and L. Shivakumar. 2004. Self-selection of auditors and audit pricing in private firms. *The Accounting Review* 79 (1): 51–72.
- Clatworthy, M., G. Makepeace, and M. Peel. 2009. Selection bias and the Big Four premium: New evidence using Heckman and matching models. *Accounting and Business Research* 39 (2): 139–166.
- Cram, D., V. Karan, and I. Stuart. 2009. Three threats to validity of choice-based and matched sample studies in accounting research. *Contemporary Accounting Research* 26 (2): 477–516.
- Craswell, A., J. Francis, and S. Taylor. 1995. Auditor brand name reputations and industry specializations. *Journal of Accounting and Economics* 20 (3): 297–322.
- Davis, L. R., D. N. Ricchiute, and G. Trompeter. 1993. Audit effort, audit fees, and the provision of nonaudit services to audit clients. *The Accounting Review* 68 (1): 135–150.
- DeAngelo, L. 1981. Auditor size and audit quality. *Journal of Accounting and Economics* 3 (3): 183–199.
- Dechow, P., R. Sloan, and A. Sweeney. 1996. Causes and consequences of earnings manipulation: An analysis of firms subject to enforcement actions by the SEC. *Contemporary Accounting Research* 13 (1): 1–36.
- DeFond, M., J. Francis, and T. Wong. 2000. Auditor industry specialization and market segmentation: Evidence from Hong Kong. *Auditing: A Journal of Practice & Theory* 19 (1): 49–66.
- Dehejia, R., and S. Wahba. 1999. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association* 94 (4): 1053–1062.
- , and ———. 2002. Propensity score-matching methods for nonexperimental causal studies. *The Review of Economics and Statistics* 84 (1): 151–161.
- . 2005. Practical propensity-score matching: A reply to Smith and Todd. *Journal of Econometrics* 125 (1–2): 355–364.
- Dopuch, N., and D. Simunic. 1980. The nature of competition in the auditing profession: a descriptive and normative view. In *Regulation and the Accounting Profession*, 34, (2): edited by J. Buckley and F. Weston, 283–289. Belmont, CA: Lifetime Learning Publications.
- Dow Jones News Service. 2002. Grant Thornton acquires Andersen partners in Houston. (June 13).
- Easton, P. 2004. PE ratios, PEG ratios, and estimating the implied expected rate of return on equity capital. *The Accounting Review* 79 (1): 73–95.
- Farber, D. 2005. Restoring trust after fraud: Does corporate governance matter? *The Accounting Review* 80 (2): 539–561.
- Ferguson, A., J. Francis, and D. Stokes. 2003. The effects of firm-wide and office-level industry expertise on audit pricing. *The Accounting Review* 78 (2): 429–448.
- Feroz, E., K. Park, and V. Pastena. 1991. The financial and market effects of the SEC's accounting and auditing enforcement releases. *Journal of Accounting Research* 29 (Supplement): 107–142.
- Financial Accounting Standards Board (FASB). 1987. *Statement of Cash Flows*. Statement of Financial Accounting Standards No. 95. Norwalk, CT: FASB.

- Francis, J. 1984. The effect of audit firm size on audit prices: A study of the Australian market. *Journal of Accounting and Economics* 6 (2): 133–151.
- , and D. Stokes. 1986. Audit prices, product differentiation, and scale economies: Further evidence from the Australian market. *Journal of Accounting Research* 24 (2): 383–393.
- , and J. Krishnan. 1999. Accounting accruals and auditor reporting conservatism. *Contemporary Accounting Research* 16 (1): 135–165.
- , E. Maydew, and H. Sparks. 1999. The role of Big 6 auditors in the credible reporting of accruals. *Auditing: A Journal of Practice & Theory* 18 (2): 17–34.
- . 2004. What do we know about audit quality? *The British Accounting Review* 36 (4): 345–368.
- , C. Lennox, and Z. Wang. 2010. Selection models in accounting research. Working paper, University of Missouri–Columbia, Hong Kong University of Science and Technology, and Nanyang Technological University.
- , and M. Yu. 2009. Big Four office size and audit quality. *The Accounting Review* 84 (5): 1521–1552.
- Frankel, R., M. Johnson, and K. Nelson. 2002. The relation between auditors' fees for nonaudit services and earnings management. *The Accounting Review* 77 (Supplement): 71–105.
- Geiger, M., and D. Rama. 2006. Audit firm size and going-concern reporting accuracy. *Accounting Horizons* 20 (1): 1–17.
- Government Accountability Office (GAO). 2003. *Public Accounting Firms. Mandated Study on Consolidation and Competition*. Washington, D.C.: GAO.
- . 2008. *Audits of Public Companies. Continued Concentration in Audit Market for Large Public Companies does not Call for Immediate Action*. Washington, D.C.: GAO.
- Guay, W., S. P. Kothari, and R. Watts. 1996. A market-based evaluation of discretionary-accrual models. *Journal of Accounting Research* 34 (Supplement): 83–115.
- Guenther, D. A., and M. Willenborg. 1999. Capital gains tax rates and the cost of capital for small business: Evidence from the IPO market. *Journal of Financial Economics* 53 (3): 385–408.
- Gul, F., K. J. Kim, and A. Qiu. 2009. Ownership concentration, foreign shareholding, audit quality, and stock price synchronicity: Evidence from China. *Journal of Financial Economics* 95 (3): 425–442.
- Guo, S., and M. Fraser. 2010. *Propensity Score Analysis: Statistical Methods and Applications*. Thousand Oaks, CA: SAGE Publications, Inc.
- Hastie, T., and R. Tibshirani. 1990. *Generalized Additive Models*. London, U.K.: Chapman and Hall.
- Hay, D. C., W. R. Knechel, and N. Wong. 2006. Audit fees: A meta-analysis of the effect of supply and demand attributes. *Contemporary Accounting Research* 23 (1): 141–191.
- Heckman, J. 1979. The sample bias as a specification error. *Econometrica* 47 (1): 153–162.
- , H. Ichimura, and P. Todd. 1997. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies* 64 (4): 605–654.
- , ———, and ———. 1998. Matching as an econometric evaluation estimator. *The Review of Economic Studies* 65 (2): 261–294.
- , and S. Navarro-Lozano. 2004. Using matching, instrumental variables, and control functions to estimate economic choice models. *The Review of Economics and Statistics* 86 (1): 30–57.
- Hribar, P., and C. Nichols. 2007. The use of unsigned earnings quality measures in tests of earnings management. *Journal of Accounting Research* 45 (5): 1017–1053.
- Krishnan, G. 2003. Audit quality and the pricing of discretionary accruals. *Auditing: A Journal of Practice & Theory* 22 (1): 109–126.
- Khurana, I., and K. Raman. 2004. Litigation risk and the financial reporting credibility of Big 4 versus non-Big 4 audits: Evidence from Anglo-American countries. *The Accounting Review* 79 (2): 473–495.
- Kothari, S. P., A. Leone, and C. Wasley. 2005. Performance matched discretionary accrual measures. *Journal of Accounting and Economics* 39 (1): 163–197.
- Lang, M., and R. Lundholm. 1996. Corporate disclosure policy and analyst behavior. *The Accounting Review* 71 (4): 467–492.
- Lennox, C. 1999. Are large auditors more accurate than small auditors? *Accounting and Business Research* 29 (3): 217–228.

- Li, K., and N. Prabhala. 2007. Self-selection models in corporate finance. In *Handbook of Corporate Finance: Empirical Corporate Finance*, edited by B. E. Eckso, 37–86. Amsterdam, The Netherlands: Elsevier Science B.V.
- Louis, H. 2005. Acquirers' abnormal returns and the non-Big 4 auditor clientele effect. *Journal of Accounting and Economics* 40 (1–3): 75–99.
- Mitton, T. 2002. A cross-firm analysis of the impact of corporate governance on the East Asian financial crisis. *Journal of Financial Economics* 64 (2): 215–241.
- O'Keefe, T. B., D. A. Simunic, and M. T. Stein. 1994. The production of audit services: Evidence from a major public accounting firm. *Journal of Accounting Research* 32 (2): 241–261.
- Palmrose, Z. V. 1988. An analysis of auditor litigation and audit service quality. *The Accounting Review* 63 (1): 55–73.
- Ramsey, J. 1969. Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series A (General)* 31 (2): 350–371.
- Rosenbaum, P., and D. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70 (1): 41–55.
- Simunic, D. 1980. The pricing of audit services: Theory and evidence. *Journal of Accounting Research* 18 (1): 161–190.
- Smart, S., and C. Zutter. 2003. Control as a motivation for under pricing: A comparison of dual and single-class IPOs. *Journal of Financial Economics* 69 (1): 85–110.
- Weber, J., and M. Willenborg. 2003. Do expert informational intermediaries add value? Evidence from auditors in microcap IPOs. *Journal of Accounting Research* 41 (4): 681–720.
- Whisenant, S., S. Sankaraguruswamy, and K. Raghunandan. 2003. Evidence on the joint determination of audit and nonaudit fees. *Journal of Accounting Research* 41 (4): 721–744.
- White, H.. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48 (4): 817–838.
- Zmijewski, M. 1984. Methodological issues related to the estimation of financial distress prediction models.. *Journal of Accounting Research* 22 (Supplement): 59–82.

The Role of the Internal Audit Function in the Disclosure of Material Weaknesses

Shu Lin

California State University, Fresno

Mina Pizzini

Southern Methodist University

Mark Vargus

Drexel University

Indranil R. Bardhan

The University of Texas at Dallas



ABSTRACT: This study investigates the role that a firm's internal audit function (IAF) plays in the disclosure of material weaknesses reported under Section 404 of the Sarbanes-Oxley Act of 2002 (U.S. Congress 2002). Using data from 214 firms, we examine the relation between material weakness (MW) disclosures and various IAF attributes and activities. Our results indicate that MW disclosures are negatively associated with the education level of the IAF and the extent to which the IAF incorporates quality assurance techniques into fieldwork, audits activities related to financial reporting, and monitors the remediation of previously identified control problems. The timing of Section 404 work and the nature of follow-up monitoring suggests that these aspects of IAF quality help prevent MWs from occurring. We find that MW disclosures are positively associated with the IAF practice of grading audit engagements and external-internal auditor coordination, suggesting that these activities increase the effectiveness of Section 404 compliance processes.

Keywords: *internal audit function; material weakness; corporate governance; internal control over financial reporting.*

Data Availability: *The Institute of Internal Auditors controls access to the Global Auditing Information Network (GAIN) survey data.*

We thank the Institute of Internal Auditors for access to the Global Auditing Information Network (GAIN) database. We thank Jeffrey Doyle, Weili Ge, and Sarah McVay for access to material weakness data reported in Doyle et al. (2007a). We also thank Larry Abbott, Barry Brian, Margaret Christ, Mark Salamasick, Douglas Ziegenfuss, Steven Kachelmeier (editor), two anonymous referees, the WWMA, and participants at the 2008 AAA Management Accounting Section Meeting, 2008 AAA Auditing Section Meeting, and the 2008 Lone Star Accounting Research Conference for their helpful comments and suggestions.

Editor's note: Accepted by Steven Kachelmeier.

Submitted: March 2008

Accepted: June 2010

Published Online: January 2011

I. INTRODUCTION

This study investigates the role that a firm's internal audit function (IAF) plays in the disclosure of material weaknesses reported under Section 404 of the Sarbanes-Oxley Act of 2002 (U.S. Congress 2002). Management is ultimately responsible for establishing and maintaining adequate internal control over financial reporting and for evaluating the effectiveness of financial reporting controls and procedures. However, "support for management in the discharge of these responsibilities is a legitimate role for internal auditors," as long as it does not impair internal audit objectivity (Institute of Internal Auditors [IIA] 2004, 3). External auditing standards have long acknowledged internal auditing as a potentially valuable resource in the financial reporting process (SAS No. 65, AICPA 1991; AS No. 2, PCAOB 2004; AS No. 5, PCAOB 2007a). Consistent with this reasoning, Bailey et al. (2003) find that companies view the IAF as a means of correcting perceived breakdowns in business reporting, internal control, and ethical behavior.

Despite the IAF's duties surrounding internal control over financial reporting (ICFR), few researchers have empirically investigated the IAF's role in the financial reporting process (Gramling et al. 2004). A notable exception is a recent study by Prawitt et al. (2009), which provides evidence that the IAF can improve reporting quality by mitigating potential weaknesses in incentive system design. Our study complements theirs, in that we examine the association between the IAF and ICFR through prevention and detection of material weaknesses. Accordingly, our study helps fill an important gap in the literature regarding the influence of the IAF on the quality of the financial reporting process.

Professional standards and prior research (AICPA 1991; IIA 2008; Prawitt et al. 2009) suggest that IAF quality encompasses specific attributes of the organizations and parties performing internal audit activities (e.g., competence of IAF personnel), as well as the nature and scope of activities performed (e.g., the extent to which IAF monitors remediation of previously identified control problems). We investigate the relation between these factors and the likelihood that a firm reports a material weakness (MW). For a material weakness to be disclosed, it must exist and be detected. It is not clear *ex ante* how our IAF quality measures will affect disclosures. For instance, more competent IAF personnel can help management establish stronger controls over financial reporting, and thus reduce the existence of control problems. Conversely, if a material weakness exists, more competent IAF personnel are more likely to detect it. We carefully consider the influence of various aspects of IAF quality on the existence, detection, and disclosure of MWs in developing our hypotheses.

We conduct our tests using data on 214 firms that provided detailed responses to the IIA's Global Auditing Information Network (GAIN) survey for 2003–2004. We identify 45 firms that disclosed at least one MW under SOX Section 404. Results indicate that the education level of IAF staff and the extent to which the IAF incorporates quality assurance techniques into fieldwork, audits activities related to financial reporting, and monitors the remediation of previously identified control problems are negatively associated with MW disclosures. The timing of Section 404 work and the nature of follow-up monitoring suggest that these aspects of IAF quality help prevent the existence of MWs. The IAF practices of grading audit engagements and coordinating with external auditors are both positively associated with MW disclosures. The positive relations suggest that these activities increase the effectiveness of Section 404 compliance processes and thereby increase the likelihood that extant MWs will be detected and disclosed. Together, our results have important implications for managers responsible for determining IAF staffing and the structure of IAF activities, external auditors who perform Section 404 work, and standard-setters who provide Section 404 guidance. Moreover, our evidence that the IAF affects the financial reporting process lends support to the requirement that NYSE-listed companies maintain an internal audit function (NYSE 2009).

This study makes several important contributions to the literature. First, we expand extant research on both internal auditing and ICFR by documenting associations between material weaknesses and various measures of IAF quality. Second, we identify specific internal auditor practices and procedures that are associated with both the detection and prevention of material weaknesses. Few archival studies directly link auditor practices to the prevention or detection of audit exceptions. Finally, we provide evidence that external auditors are more likely to detect material weaknesses when they coordinate their efforts with the IAF. While a significant branch of the internal auditing literature focuses on the relation between external auditors and the IAF, no study of which we are aware has examined whether IAF involvement in external audits increases the effectiveness of external audits.

Section II provides background information and reviews relevant literature, Section III describes our hypotheses, and Section IV describes the sample and model. Results are discussed in Section V, with concluding remarks in Section VI.

II. BACKGROUND AND LITERATURE REVIEW

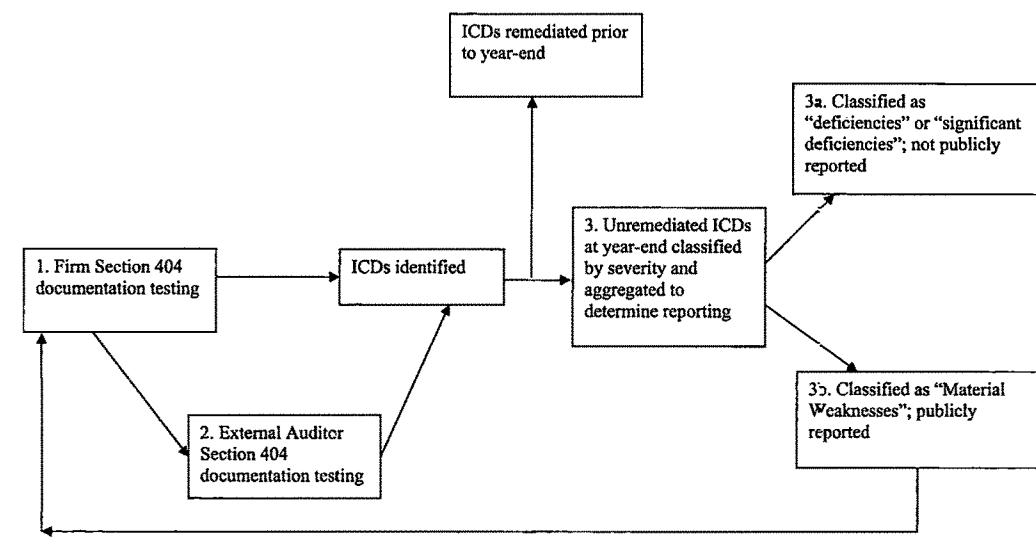
Material Weakness Reporting

A material weakness is a deficiency, or combination of deficiencies, that result in a reasonable possibility that a company's controls will fail to prevent or detect a material misstatement of an account balance or disclosure (AS No. 5, PCAOB 2007a).¹ Section 404 of SOX mandates that managers evaluate internal control over financial reporting (ICFR) and present the results of their evaluation in financial statements filed on Form 10-K and Form 10-Q. The regulation, which became effective for accelerated filers for year-end dates beginning November 15, 2004, also requires external auditors to annually assess and state an opinion on ICFR. Accordingly, both management and external auditors are responsible for ensuring that material weaknesses are detected and disclosed under Section 404. Prior to the implementation of Section 404, Section 302 required management to evaluate and report on the effectiveness of ICFR; however, the external auditor was not required to opine on ICFR (U.S. Congress 2002). We exclude Section 302 material weaknesses from our analysis because these disclosures were subject to less regulation and allowed greater management discretion than were material weaknesses reported under SOX 404 (Ashbaugh-Skaife et al. 2007; Hoitash et al. 2009).

For a material weakness to be reported, it must exist, it must be detected, and it must be disclosed (Ashbaugh-Skaife et al. 2007). Figure 1 summarizes the sequence of Section 404 processes in a given year (Bedard and Graham 2011). Step 1 represents the firm's documentation and testing performed in support of management's evaluation of internal controls. Step 2 represents the external auditor's documentation and testing, which *must follow* the firm's work in a given area/account. At the end of the year, internal control deficiencies (ICDs) that have not been remediated are classified by severity and aggregated to determine whether they constitute a material weakness and, hence, must be publicly disclosed. The PCAOB and SEC direct external auditors and managers to evaluate the severity of each control deficiency to determine whether the deficiencies, individually or in combination, constitute a material weakness as of the date of management's assessment (AS No. 5, PCAOB 2007a; SEC 2007). Severity depends upon whether there is a reasonable possibility that the company's controls will fail to prevent or detect a misstatement and the magnitude of the potential misstatement. Multiple deficiencies that affect the same account or

¹ Although AS No. 2 was in effect during our sample period, we refer to AS No. 5, which supersedes AS No. 2, because we do not expect the differences in the two standards to affect our predictions regarding IAF quality (PCAOB 2004, 2007a). Two key differences are that AS No. 5 eliminates the redundant requirement that external auditors opine on both internal controls and on management's assessment of internal controls and replaces the words "more than remote" with the words "reasonable possibility" for defining what constitutes a material weakness.

FIGURE 1
Sequence of Section 404 Compliance Processes



disclosure may, in combination, constitute a material weakness, even though individually the deficiencies are less severe.

Gramling et al. (2004, 196) suggest that management's increased accountability for ICFR under SOX implies an expanded role for the IAF. Consistent with this reasoning, the IIA provides specific guidance regarding how the IAF can support management's SOX compliance (IIA 2004). The IIA advocates that the IAF independently evaluate management's testing and assessment processes, and management's basis for their assertions regarding the adequacy of internal controls. If control gaps are identified, then internal auditing should assess management's plans for correcting them and perform follow-up reviews. The IAF can also perform effectiveness testing for reliance by external auditors. Finally, the IIA advocates that the IAF act as coordinator between management and the external auditors and ensure that the results of ongoing internal audit activities are disclosed.

Prior Research on Auditor Quality

Practitioners and academics alike generally contend that the effectiveness of internal controls is increasing in IAF quality. However, direct empirical evidence of this relation is limited (e.g., AICPA 1991; Kinney 2000; Bailey et al. 2003; Gramling et al. 2004; PCAOB 2007a; Prawitt et al. 2009).² In an experimental setting, Schneider and Wilner (1990) find that, under certain conditions, internal auditing deters financial reporting irregularities. In an archival study of Australian firms, Davidson et al. (2005) find no significant relation between the presence (versus absence) of an IAF and earnings management. Prawitt et al. (2009) investigate the relation between IAF quality and earnings management using the GAIN database. Guided by external auditing standards, they develop a composite measure of IAF quality from proxies for IAF competence, IAF

² See Gramling et al. (2004) for a review of the literature on internal auditing.

objectivity, the amount of financial reporting work the IAF performs, and IAF size. They find evidence that IAF quality is associated with a lower level of earnings management. Their results suggest that higher-quality IAFs are more likely to deter managers from manipulating earnings and/or detect such manipulations and ensure that they are corrected prior to the issuance of financial statements.

Two recent studies of ICDs identified under Sections 302 and 404 suggest that the quality of the auditors performing Section 302 and Section 404 work is positively associated with the detection and disclosure of control problems. Using proprietary data on small accelerated filers, Bedard and Graham (2011) find that firms that use outside consultants for Section 404 compliance report higher levels of ICD severity. They posit that outside consultants have greater expertise in control assessments that improves the effectiveness of Section 404 processes and increases the likelihood of ICD detection and the appropriateness of ICD classification. In a study of non-accelerated filers, Bedard et al. (2009) find that the disclosure of Section 302 ICDs increases with the experience of the external audit firm in conducting Section 404 work, suggesting that audit firms leverage the knowledge gained from SOX 404 audits in performing SOX 302 work. They also find that more Section 302 MWs are reported in the fourth quarter, when there is greater external auditor involvement relative to the first three quarters. They contend that external audit firms' expertise, experience, and compliance processes for Section 302 are superior to that of client firms, and thus lead to increased detection and disclosure of ICDs at year-end.³ While these studies provide useful archival evidence on the relation between auditor characteristics and the detection and disclosure of ICDs, neither directly addresses internal auditor quality.

Research on the effects of third-party monitoring on management reporting is also relevant to internal auditing (Prawitt et al. 2009). Hoitash et al. (2009) find that greater accounting and financial supervisory experience on the audit committee is associated with a lower likelihood of Section 404 MW disclosures. Studies of management forecasts and communications show that management is less biased when their bias is likely to be perceived by others (Schwartz and Young 2002; Rogers and Stocken 2005). Brown and Pinello (2007) provide evidence that the annual reporting process, which calls for an external audit, mitigates earnings management in year-end reports relative to that in interim reports. They attribute this result, at least in part, to increased scrutiny of year-end reports. Accordingly, if the internal auditor function is viewed as a third party that monitors managers' actions on a year-round basis, then improvements in IAF quality should strengthen deterrence and detection mechanisms.

III. HYPOTHESIS DEVELOPMENT

The overall relation between IAF quality and MW disclosures depends on how various IAF attributes and activities affect the existence, detection, and disclosure of MWs. If the effectiveness of ICFR is increasing in IAF quality, then IAF quality should be negatively associated with the existence of control problems and positively associated with both the detection and disclosure of extant control problems. Greater IAF quality deters managers from taking actions that compromise controls and encourages managers to put strong controls in place. Albrecht and Albrecht (2004) note that an effective control structure is probably the single most important step to eliminate (or minimize) the opportunity to commit fraud. A more capable IAF is also more likely to detect and correct minor control problems before they become severe enough to be considered material weaknesses. Decreasing the existence of control deficiencies unambiguously reduces the likelihood that a material weakness is reported.

³ Competing disclosure incentives of management and the external auditor also explain the increased disclosure rate (Ashbaugh-Skaife et al. 2007; Asare et al. 2007).

If a material weakness already exists, then we expect that the improved detection capabilities of higher quality IAFs will increase the likelihood of disclosure. When a material weakness exists, the relation between IAF detection capability and disclosure depends upon (1) the likelihood that the external auditors will identify the problem during their year-end audit without the help of the IAF,⁴ and (2) management's ability to correct the weakness prior to year-end, which is affected by the timing of IAF Section 404 procedures. If we assume the external auditors will not discover an existing MW, then IAF detection directly increases the likelihood of disclosure.⁵ In this scenario, management can avoid disclosure only if they can correct the problem prior to year-end.⁶ If we assume that the external auditors will detect all extant MWs at year-end, regardless of IAF activities, then IAF detection of a MW prior to year-end can potentially decrease the likelihood of disclosure by providing management with an opportunity to remediate the control weakness.

However, remediation requires that the IAF detect the problem early enough in the fiscal year for management to rectify the problem and that management deems the problem severe enough to take action. Ettredge et al. (2006) and PCAOB reviews of AS No. 2 (PCAOB 2005, 2007b) suggest that firms' Section 404 procedures were performed near or at year-end during the time period of our sample (2004–2006), leaving management little time for remediation. Moreover, Bedard and Graham (2011) report that management tends to underclassify the severity of control problems relative to external auditors, and thus may not recognize that a control problem is serious enough to warrant corrective action. Given this evidence, we conclude that it is unlikely that management will be able to successfully remediate the control problems in a timely manner. Accordingly, we expect that better IAF detection capabilities will lead to more MW disclosures, although we cannot rule out the possibility that enhanced IAF detection capabilities can reduce MW disclosures.

We draw on Prawitt et al. (2009) and professional guidance to develop measures of IAF quality (AICPA 1991; IIA 2008). These suggest that IAF quality measures encompass (1) competence, (2) objectivity, (3) relative investment in the IAF, and (4) the nature and scope of IAF activities. We group the first three measures together as IAF attributes because they address the characteristics of organizations performing internal audit activities, and thus are covered by the Attribute Standards of the *International Standards for the Professional Practice of Internal Auditing* (IIA 2008). The last construct, the nature and scope of IAF activities, captures IAF practices that are guided by Performance Standards (IIA 2008).⁷

IAF Quality Attributes

IIA Attribute Standards stipulate that internal auditors possess the knowledge, skills, and other competencies needed to effectively carry out their responsibilities (IIA 2008). To prevent and detect internal control irregularities, internal auditors must have a thorough understanding of company operations, processes, and procedures, and they must be able to design and implement tests to determine whether processes and procedures are working as intended (Clark et al. 1980).

⁴ The literature is only beginning to address this complex issue of which party (firm or external auditor) is more likely to detect control deficiencies (Bedard et al. 2009; Bedard and Graham 2011). Therefore, we do not attempt to assess the likelihood that the external auditors would fail to detect a MW had it not been identified by the internal auditors.

⁵ IIA guidance specifies that the IAF report all significant risk exposures and control issues to senior management and the board. Senior management is obligated under Section 404 to disclose a material weakness. It is unlikely that both management and the audit committee would knowingly violate SOX by failing to disclose a significant control problem to the external auditors.

⁶ According to the SEC, "if management's evaluation process identifies material weaknesses, but all material weaknesses are remediated by the end of the fiscal year, management may conclude that internal control over financial reporting is effective at the end of the fiscal year" (SEC 2007, 10, footnote 20).

⁷ We use the most recent version of the IIA Standards, issued October 2008, to guide our choice of quality measures. This version is very similar to earlier versions that were in place at the time the GAIN survey was conducted.

External auditing standards state that external auditors should consider professional certification, professional experience, and training in evaluating internal auditor competence (SAS No. 65, AICPA 1991). Consistent with this statement, prior research suggests that external auditors' evaluations of internal auditor competence are based upon professional certifications (Brown 1983) and the experience of IAF personnel (Messier and Schneider 1988).

IIA Standards require internal auditors to be independent and objective in performing their work (IIA 2008). An objective IAF is less likely to be influenced by management in evaluating controls and reporting internal control problems to the audit committee. Professional governance guidelines and standards suggest that the reporting relationship between the Chief Audit Executive (CAE) and audit committee is a key determinant of internal auditor objectivity (Gramling et al. 2004). Consistent with this premise, Bedard and Graham (2011) find that ICD severity levels are higher when the party performing the firm's Section 404 work reports directly to the audit committee rather than to management.

Practitioners and academics generally contend that management can improve IAF quality by increasing resources allocated to the IAF (e.g., Gramling et al. 2004; Ge and McVay 2005). Greater IAF resources enable management to hire and retain more competent personnel and improve the effectiveness of IAF consulting and assurance activities. In a descriptive analysis of MW disclosures, Ge and McVay (2005) find that poor internal control is usually related to an insufficient commitment of resources to accounting controls. Studies of the economic determinants of ICD disclosures find that firms with fewer resources available for internal controls are more likely to disclose ICDs (Ashbaugh-Skaife et al. 2007; Doyle et al. 2007a). Recent external auditing studies suggest that the greater resources available to large and mid-tier audit firms, relative to small audit firms, enable them to develop more effective SOX compliance processes that lead to greater ICD detection and disclosure (Bedard et al. 2009; Bedard and Graham 2011). Thus, we expect firms with greater IAF investment will implement stronger Section 404 procedures in support of management's evaluation of ICFR.

It is not clear *ex ante* how IAF quality attributes will affect MW disclosures. Competence, objectivity, and investment enhance quality and, thus, should reduce the likelihood that material weaknesses exist. Conversely, if a material weakness does exist, a higher quality IAF will be more likely to detect it. Given the bi-directional implications for material weakness disclosures, we present a nondirectional hypothesis for the three quality attributes.

H1: The likelihood that a firm reports a material weakness is significantly associated with IAF competence, IAF objectivity, and IAF investment.

Nature and Scope of IAF Activities

External auditing standards contend that external auditors evaluate the nature, timing, and extent of IAF fieldwork in audit planning and determining whether to rely on the work of internal auditors. AICPA and IIA guidance suggest that the following factors are relevant to the financial reporting process: (1) use of fieldwork quality assurance techniques, (2) inclusion of financial reporting processes in audit scope, (3) communication of grades or summary opinions on control effectiveness, (4) follow-up of previously identified control problems, and (5) coordination with external auditors (AICPA 1991; IIA 2008).

The quality of audit fieldwork performed is critical to the detection of internal control deficiencies. Schneider (1984, 1985) and Brown and Karan (1986) find that external auditors place more emphasis on the quality of work performance than either competence or objectivity in evaluating the IAF. Quality assurance (QA) techniques help ensure that IAF fieldwork is effective and appropriate. QA activities include direct supervision, independent working paper review, solicitation of audit client feedback, peer review by fellow staff members, and the use of working paper checklists.

As Prawitt et al. (2009, 1261) note, firms have broad latitude in setting the scope of the IAF, and its focus can vary widely across organizations. IAF responsibilities generally encompass a range of control activities not directly related to financial reporting, such as operational audits, systems audits, and internal consulting projects. Since ICFR forms only a subset of a firm's overall internal controls, external auditors must evaluate the extent to which IAF activities are relevant to the financial statement audit in determining whether they can rely on IAF work (SAS No. 65; AICPA 1991). We expect that greater IAF attention to financial reporting will affect both the existence and detection of material weaknesses.

Internal auditing standards direct internal auditors to document and communicate engagement results to management, and, where appropriate, include their overall opinion (IIA 2008). A grade or rating is a succinct means of conveying an opinion on the risk posed by the unit or functional area audited, and the IIA provides specific guidance on the practice of rating internal controls (IIA 2009a). A recent study by PricewaterhouseCoopers (PwC 2006) reports an increase in the prevalence of grading and concludes that grading is now generally considered to be a best practice. Grading motivates managers to put adequate controls in place to avoid a poor grade and facilitates design and implementation of monitoring systems. Both help prevent the existence of control problems. Grading also promotes rapid assessment of control risks by management, audit committees, and external auditors (PwC 2006), and thus can help management and external auditors implement the risk-based approach to Section 404 compliance recommended by professional guidance (PCAOB 2007a; SEC 2007). As an attention-directing tool, grading improves an auditor's assessment of the risk of financial misstatements and thereby facilitates appropriate allocation of audit resources to the evaluation and testing of relevant company-level controls (Hogan and Wilkins 2008; PCAOB 2007a; Wright and Ashton 1989). Accordingly, we expect IAF grading to improve the effectiveness of both internal and external auditors' Section 404 audit procedures, thus increasing the likelihood that extant material weaknesses are detected.

As with the IAF quality attributes, it is not clear *ex ante* how using QA techniques, focusing on financial reporting controls, and grading internal controls will affect the likelihood that a material weakness is reported. These monitoring activities can prompt managers to take preventative action in anticipation of the IAF's rigorous review (Brown and Pinello 2007). Conversely, such activities increase the likelihood that the IAF discovers and discloses any existing control problems. Further, IAF grading can increase the effectiveness of both internal and external Section 404 compliance processes, leading to greater MW detection. Due to the bi-directional implications for material weakness disclosures, we test a nondirectional hypothesis.

H2: The likelihood that a firm reports a material weakness is associated with: (1) the extent to which the IAF uses quality assurance techniques, (2) the extent to which IAF activities address financial reporting activities, and (3) whether the IAF grades audit engagements.

IIA performance standards and SOX guidance require the Chief Audit Executive (CAE) to establish and maintain a system to monitor the disposition of previously identified control problems (IIA 2008, 2002). Follow-up procedures can reduce the likelihood that a material weakness exists at year-end and, hence, must be publicly disclosed. First, follow-up procedures provide the impetus for management to correct less severe control problems, thereby preventing them from becoming material weaknesses. Second, if the firm or its external auditors detect an existing material weakness at any time prior to year-end, then management can avoid disclosing it if they resolve the problem prior to year-end. Hence, IAF follow-up procedures can prompt management to remediate identified weaknesses in a timely manner. Accordingly, we predict that follow-up procedures will be negatively associated with MW disclosures.

H3: Firms whose IAFs follow-up on control problems are less likely to report material weaknesses.

Both internal and external auditing standards encourage auditors to coordinate efforts related to an integrated audit under AS No. 5 (PCAOB 2007a; IIA 2008). Methods of coordination include conducting joint risk or planning sessions, performing audits of specific processes or locations, and loaning IAF staff to the external auditor. As firms' "in-house" experts on internal control, internal auditors possess company-specific knowledge about controls, operations, and the financial reporting process that can aid external auditors in implementing the top-down, risk-based approach recommended in AS No. 5 (PCAOB 2007a). Coordination should improve an external auditor's assessment of the risk of misstatements, thus facilitating appropriate allocation of audit resources to evaluation and testing of relevant controls (Hogan and Wilkins 2008; PCAOB 2007a). Reliance on the work performed by the IAF, both independently and under the direction of the external auditors, can improve the effectiveness of Section 404 documentation and testing.⁸ According to the PCAOB (2007b, 8–9), "an auditor who appropriately uses the work of others can achieve the objectives of the audit while not duplicating effort in lower-risk areas, and also is better able to focus his or her own efforts on higher-risk controls." Moreover, to the extent that coordination reduces time pressure on the external auditor, audit effectiveness should increase with coordination (McDaniel 1990). Given the value the IAF can add to the external auditor's Section 404 processes, we propose that internal-external auditor coordination will enhance the overall effectiveness of the external auditor's Section 404 process, thereby leading to greater detection and disclosure of MWs.⁹

Our prediction assumes that most of the external auditor's Section 404 documentation and testing takes place at year-end, giving management little opportunity to correct any control problems detected by the external auditors. While we recognize that coordination can and does occur throughout the year, we expect that much of the benefit of that coordination is realized at year-end, when external auditors conduct their most rigorous reviews of the financial statements and the reporting process (Frankel et al. 2002; Brown and Pinello 2007; Bedard et al. 2009). In particular, the rapid enactment of Section 404 combined with delays in managements' completion of SOX procedures left external auditors little choice but to perform most Section 404 work at year-end for the time period of our sample of MWs (PCAOB 2005; Ettredge et al. 2006). The benefits of IAF coordination and the year-end timing of external auditors' Section 404 processes lead to the following hypothesis:

H4: Firms in which IAFs coordinate with external auditors are more likely to report material weaknesses.

IV. RESEARCH DESIGN

Data and Sample Selection

Data for this study come from multiple sources. We use firm-level data collected by the IIA through their GAIN survey for the years 2003 and 2004. The GAIN database consists of CAEs' responses to a comprehensive survey designed to measure various aspects of an organization's

⁸ The IAF must be of sufficient quality, i.e., it must be objective and competent for external auditors to rely on their work. Thus, even though the IAF are employees of the firm being audited, we expect that the IAF will be objective in assessing controls and reporting control problems to the external auditors.

⁹ External auditor reliance on the IAF is increasing in IAF quality (e.g., Schneider 1984, 1985; Maletta 1993), and as previously discussed, it is not *a priori* clear how IAF quality affects the likelihood of MW disclosure. Our tests of the relation between IAF coordination and MW disclosure include explicit measures of IAF quality, and thereby control for the effect of IAF quality on disclosure.

internal audit activities. The annual survey captures information on several topics, including internal audit department costs, oversight and reporting responsibilities, audit committee, audit life-cycles, and financial measures.¹⁰ Next, we collect firm data on MW disclosures from EDGAR, firm financial data from Compustat, stock return data from CRSP, audit fee and restatement data from Audit Analytics, and institutional ownership data from CDA/Spectrum.

Our initial sample contained 1,356 responses from 935 GAIN survey respondents collected in years 2003 and 2004. Since firm names are not reported in the survey data, we employ a matching algorithm to identify firms based on reported SIC code, total assets, revenues, and number of employees. Next, we eliminate: (1) 687 firms for which data were unavailable on Compustat, (2) six firms with American Depositary Receipts or missing CRSP data, (3) six firms with missing 10-K data after August 2002, (4) six firms missing stock information from both Compustat and CRSP, and (5) eight firms missing necessary GAIN data, resulting in a sample of 329 firm-year observations from 222 firms.¹¹ For firms that report survey data for both 2003 and 2004, we use only the data from the 2004 survey year, which yields our final sample of 222 firms. Finally, we eliminate five firms that only report material weaknesses under SOX Section 302, and three firms that are non-accelerated filers and therefore not subject to SOX Section 404 reporting requirements.¹² The final sample consists of 214 firms (Table 1).

IAF Quality as a Determinant of Material Weaknesses

We model the probability that a firm reports a material weakness as a function of IAF attributes, the nature and scope of IAF activities, and a set of control variables (Equation (1)). We estimate this model using a logit regression.

$$\begin{aligned}
 Prob(MW = 1) = & \beta_0 + \beta_1 EXPERIENCE + \beta_2 EDUCATION + \beta_3 CERTIFICATION \\
 & + \beta_4 TRAINING + \beta_5 CAEAC + \beta_6 CAEOFFICER + \beta_7 IASIZE \\
 & + \beta_8 FIELDWORKQA + \beta_9 IAGRADE + \beta_{10} FINANCIALFOCUS \\
 & + \beta_{11} FOLLOWUP + \beta_{12} COORDINATION + \beta_{13} SEGMENTS \\
 & + \beta_{14} FOREIGNTRANSACTIONS + \beta_{15} M\&A + \beta_{16} RESTRUCTURE \\
 & + \beta_{17} SALESGROWTH + \beta_{18} INVENTORY + \beta_{19} MARKETVALUE + \beta_{20} LOSS \\
 & + \beta_{21} CFO + \beta_{22} SHUMWAY + \beta_{23} AGE + \beta_{24} AUDITORSPECIALIST \\
 & + \beta_{25} BLUERIBBONAC + \beta_{26} ISO9000 + \beta_{27} RESTATEMENT \\
 & + \beta_{28} AUDITORRESIGN + \beta_{29} INSTITUTIONALOWNERSHIP \\
 & + \beta_{30} REGULATEDINDUSTRY + \varepsilon
 \end{aligned} \tag{1}$$

¹⁰ The data were subject to various validation checks, including validation measures built into the questionnaire and manual procedures and reasonableness tests applied after the data had been collected. The GAIN database covers a wide range of institutions including publicly traded companies, private companies, educational institutions, divisions within companies, and governmental institutions. More information can be found at the GAIN website: <http://www.theiaa.org/guidance/benchmarking/gain/>.

¹¹ We eliminate firms with missing values for *IASIZE*, *EXPERIENCE*, *EDUCATION*, and *CERTIFICATIONS*, and set missing values to zero for *TRAINING*, *OBJECTIVITY*, *FIELDWORK QA*, *FOLLOWUP*, *IAGRADE*, *FINANCIALFOCUS*, and *COORDINATION*. There are no missing values for *CAEOFFICER*, *BLUERIBBONAC*, and *ISO9000*. In a robustness check, we remove the firms with missing values in any IAF variables and obtain similar results.

¹² Inclusion of the five firms that report material weaknesses under Section 302 does not materially affect our results or conclusions. In untabulated robustness tests, the coefficient on *FINANCIALFOCUS* retains its hypothesized sign, but becomes insignificant at the 10 percent level; coefficients on the other IAF quality measures have the same signs and similar significance levels.

TABLE 1
Sample Selection and Distribution by Year

Panel A: Sample Selection

	<u>Firm-Year Observations</u>	<u>Firms</u>
Firm-year observations obtained from GAIN database for survey years 2003 and 2004 ^a	1,356	935
Less:		
Those not covered by Compustat	(990)	(687)
American Depositary Receipts	(10)	(6)
Those missing SEC filings after August 2002	(6)	(6)
Those missing stock information from both Compustat and CRSP	(10)	(6)
Those missing IAF data from GAIN	(11)	(8)
	329	222
Elimination of 2003 firm-years if data from 2004 survey-year data are available ^b	(107)	—
Elimination of firms that only report material weakness(es) under SOX Section 302	(5)	(5)
Elimination of firms that are non-accelerated filers (with market value less than \$75 million) and therefore are not subject to SOX Section 404	(3)	(3)
	214	214

Panel B: Distribution of IAF Data by Fiscal Years

<u>Fiscal Year</u>	<u>n</u>	<u>%</u>
2000	6	2.80
2001	42	19.63
2002	57	26.64
2003	97	45.32
2004	12	5.61
Total	214	100.00

^a A survey collected in 2003 or 2004 may describe earlier fiscal years. For example, survey information collected in 2003 may pertain to 2002 or 2001.

^b We retained the survey data that pre-dates the year in which a material weakness is first disclosed. In the final sample, all but two firms' IAF data describe firm characteristics that had been in existence at least one year prior to the year in which a material weakness had been disclosed. The empirical results remain similar after excluding the two firms from the sample.

where *MW* is an indicator variable that is equal to 1 if the firm disclosed a material weakness in internal control, and 0 otherwise. All model variables are defined in Table 2.

Material Weakness Firms

To construct the dependent variable, we examine 10-Ks and 10-Qs for the 214 sample firms in the EDGAR database and identify MW disclosures during the period November 15, 2004 to December 2006.¹³ We identify 47 firms that disclosed at least one material weakness during this time frame. We remove two firms due to missing data, leaving 45 firms that report at least one material weakness.

¹³ November 15, 2004 coincides with the date that Section 404 became operational for accelerated filers.

TABLE 2
Variable Definitions

Internal Control over Financial Reporting

<i>MW</i>	= An indicator variable that equals 1 if the firm disclosed a material weakness under Section 302 or Section 404, and 0 otherwise.
IAF Attributes^a	
<i>EXPERIENCE</i>	= Average number of years of auditing experience (internal and external) of the audit staff (B6a and B6b).
<i>EDUCATION</i>	= Average of the number of years of undergraduate and graduate education of the audit staff, based on highest degree achieved. Associate, Bachelor, Master, and Ph.D. degrees are assumed to take 2, 4, 6, and 8 years of study, respectively (B5).
<i>CERTIFICATIONS</i>	= Percentage of professional staff members with one or more audit certifications (B8).
<i>TRAINING</i>	= Annual hours of training per internal auditor (B15b).
<i>CAEAC</i>	= An average of eight 0-1 survey items that measures the amount of relevant internal control information the CAE reviews with the audit committee. Four items address the control environment: risk assessment system, overall assessment of corporate control environment, assessment of control environment by major subsidiary of operating entity, coordination of internal auditing with external auditor's plan. Four items address the IAF: significant findings, audits performed, fraud-conflicts of interest; and results of monitoring programs concerning compliance with law, code of conduct and ethics (C9, C10).
<i>CAEOFFICER</i>	= An indicator variable that equals 1 if the CAE position is an officer of the firm (E7).
<i>IASIZE</i>	= The total annual operating costs of the IAF (B1g) divided by the firm assets (A2), multiplied by 100.
IAF Activities^a	
<i>FIELDWORKQA</i>	= An average of seven survey items coded from 0 to 2 (0: Never, 1: Occasionally, and 2: Regularly) that measures the extent to which various quality assurance (QA) techniques are used in fieldwork. The QA techniques are: direct supervision; independent working paper review; audit client feedback; peer review by fellow staff members, working paper checklists, and ticklists; management participation; and other (H4).
<i>IAGRADE</i>	= An indicator variable that equals 1 if the final internal audit report includes a grade or score as determined by the results of the audit (H12).
<i>FINANCIALFOCUS</i>	= An average of six survey items that measure how frequently the IAF performs audits of various financial activities (0: Never, 1: Occasionally, and 2: Regularly). The survey classifies the following five activities as financial: adequacy of internal accounting controls; accuracy, reliability, and completeness of financial records; usefulness of financial reporting for management control and decision making; impact of changes in accounting rules or regulations; interim quarterly financial results reported externally (G4). We also include the frequency of GAAP compliance audits (G3e).
<i>FOLLOWUP</i>	= An indicator variable that equals 1 if there is a formal follow-up procedure to test the implementation of corrective action (H11).
<i>COORDINATION</i>	= An indicator variable that equals 1 if internal audit coordinated audit services the external auditors (F9a).

(continued on next page)

TABLE 2 (continued)

Internal Control over Financial Reporting**Control Variables**

<i>SEGMENTS</i>	= Natural log of the sum of the number of operating and geographic segments in 2003 reported by Compustat Segment file.
<i>FOREIGNTRANSACTIONS</i>	= An indicator variable that equals 1 if the firm reports a non-zero value for foreign currency translations (#150) in 2003.
<i>M&A</i>	= Proportion of years from 2001 to 2005 that a firm was involved in a merger or acquisition (Compustat AFNT #1).
<i>RESTRUCTURE</i>	= Proportion of years from 2001 to 2005 that a firm was involved in restructuring. We identify a firm undergoing a restructuring if any of the following Compustat data items are non-zero: #376, #377, #378, or #379.
<i>SALESGROWTH</i>	= Proportion of years from 2001 to 2005 that a firm's annual industry-adjusted sales growth (#12) falls into the top quintile.
<i>INVENTORY</i>	= Average inventory to total assets (#3/#6) from 2001 to 2005.
<i>MARKETVALUE</i>	= Natural log of the market value of equity (#199 × #25) in 2003.
<i>LOSS</i>	= Proportion of years from 2001 to 2005 that a firm reports negative earnings before extraordinary items (#18).
<i>CFO</i>	= Average cash flows from operations to total assets (#308/#6) from 2001 to 2005.
<i>SHUMWAY</i>	= The probability of bankruptcy as predicted by Shumway's (2001) default hazard model.
<i>AGE</i>	= Natural log of the number of years the firm has Compustat data since 1980.
<i>AUDITOR SPECIALIST</i>	= An indicator variable that equals 1 if the firm was audited by an industry specialist auditor. We define industry specialist as an auditor that collects the greatest percentage of its audit fees in the client's industry.
<i>BLUERIBBONAC</i>	= A score with a range of 0 to 1 measuring the extent to which the firm has implemented the Blue Ribbon Audit Committee recommendations. The score is an average of ten Blue Ribbon Committee recommendations coded 0/1: all members meet new criteria of independence; at least three independent outside directors; all members possess core skills including finance literacy; committee's charter is reassessed annually; proxy report states committee fulfills its charter; committee is accountable for auditor relations; outside auditors disclose all consulting assignments; auditors discuss adequacy of company accounting; 10-K's MD&A discloses financial statement reviews and discussions; and auditors review quarterly reports and 10-Q before release.
<i>ISO9000</i>	= An indicator variable that equals 1 if the firm is involved in ISO 9000.
<i>RESTATEMENT</i>	= An indicator variable that equals 1 if the firm had a restatement from 2001 to 2005.
<i>AUDITORRESIGN</i>	= An indicator variable that equals 1 if the firm experienced an auditor resignation during 2001 to 2005.
<i>INSTITUTIONAL OWNERSHIP</i>	= The percentage of institutional ownership as of December 31, 2003.
<i>REGULATED INDUSTRY</i>	= An indicator variable that equals 1 if the firm is in financial services or utility industry.

^a Letters and numbers in parentheses for IAF attributes and activities refer to item codes in 2004 GAIN Survey.

IAF Attributes

IAF quality attributes include the competence and objectivity of internal audit personnel and the amount firms invest in the IAF. In measuring these attributes, we generally follow professional guidance and Prawitt et al. (2009), who also use the GAIN data to measure IAF quality. To measure IAF competence, we use four variables, *EXPERIENCE*, *EDUCATION*, *CERTIFICATION*, and *TRAINING* (SAS No. 65, AICPA 1991; Prawitt et al. 2009). *EXPERIENCE* is defined as the average number of years of internal and external auditing experience of the audit staff; *EDUCATION* is the average number of years of undergraduate and graduate education; *CERTIFICATION* is the percentage of professional staff members with one or more audit certifications; and *TRAINING* is the average annual hours of training per staff member.¹⁴ IIA standards and guidelines for objectivity recommend that the CAE regularly communicate the results of ongoing internal audit activities to the audit committee (Practice Advisory 1110-1, IIA 2009b, 1), and that the CAE report to a level within the organization “that allows the internal audit activity to fulfill its responsibilities” (Attribute Standard 1110, IIA 2008, 2). We capture CAE-audit committee communication with the variable *CAEAC*, which measures the amount of control-related information the CAE reviews with the audit committee. Examples of such information include assessments of the control environment in total and by major subsidiary, areas audited, significant findings, fraud, and ethics compliance. A second proxy for objectivity, *CAEOFFICER*, equals 1 if the CAE position is an officer of the firm, and 0 otherwise. We expect that a CAE who is an officer will have greater support from the board and senior management, which in turn should assist the IAF in performing work free from interference (Practice Advisory 1110-1, IIA 2009b).¹⁵ We measure IAF investment, *IASIZE*, as the total IAF annual operating costs divided by total assets, based on a measure developed by Prawitt et al. (2009).¹⁶

Nature and Scope of IAF Activities

The nature and scope of IAF activities investigated includes: (1) use of quality assurance techniques in fieldwork, (2) inclusion of financial reporting processes in audit scope, (3) inclusion of grades or summary opinions on control effectiveness in audit reports, (4) follow-up of previously identified control problems, and (5) coordination with external auditors. We measure the extent to which the IAF uses quality assurance techniques, *FIELDWORKQA*, with a summative score of variables that measure the frequency with which various QA techniques are used. *FINANCIALFOCUS* is an index compiled from several items that measure the extent to which the IAF audits various activities related to financial reporting.¹⁷ The indicator variable *IAGRADE* equals 1 if the final internal audit report includes a “grade” or “score” for areas reviewed. The indicator variable *FOLLOWUP* equals 1 if internal auditors formally monitor the resolution of

¹⁴ Our measures for *EXPERIENCE* and *CERTIFICATIONS* differ slightly from Prawitt et al. (2009), while our training measure is the same. Prawitt et al. (2009) include only internal audit experience in their experience measure; we include both internal and external audit experience because external auditing experience is also valuable in conducting internal audits. Prawitt et al. (2009) calculate *CERTIFICATIONS* with two different survey items that measure separately the number of CIA and CPA designations. This approach double-counts individuals who have both designations. To avoid double-counting, we use a single survey question that asks for the “total number of professional staff with one or more audit certifications.”

¹⁵ Prawitt et al. (2009) measure objectivity with a single dichotomous variable indicating whether the CAE reports to the audit committee. Robustness tests using this variable give materially similar results in our sample.

¹⁶ Prawitt et al. (2009) also scale IAF operating costs by assets. However, they then subtract the average level of industry investment (computed using all GAIN firms) from this amount and convert this measure to a dichotomous variable by assigning a value of 1 to firms whose investment equals or exceeds the average level of investment for that firm’s particular industry. We use Prawitt et al.’s (2009) dichotomous measure in robustness tests and it does not affect the signs or significance of our IAF measures.

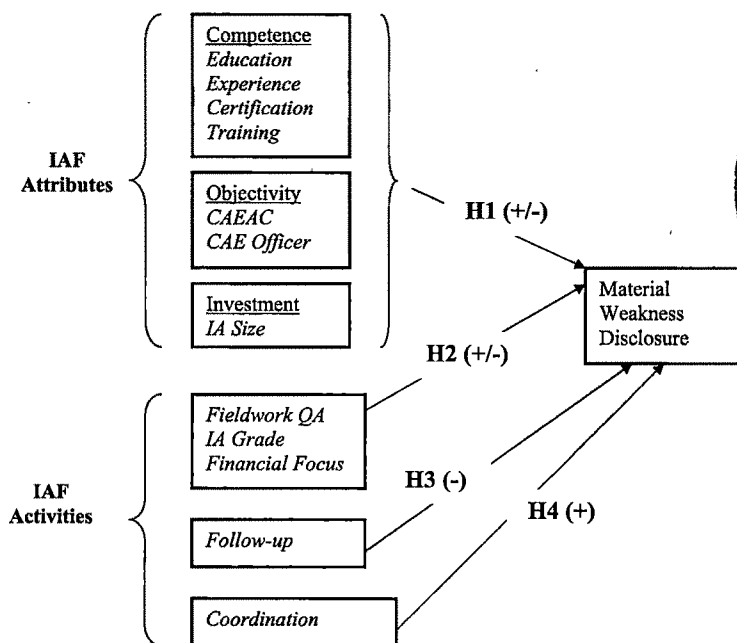
¹⁷ Prawitt et al. (2009) measure *FINANCIALFOCUS* with a survey item that gives the percentage of time the IAF spends performing financial audit work. In our sample, there are too many missing values (78 out of 214) for this item.

previously identified control problems, and 0 otherwise. The indicator variable *COORDINATION* is coded 1 if internal auditors coordinate services with external auditors, and 0 otherwise. The survey cites the following coordination methods: loaning staff to external auditors; performing complete or partial audits of specific locations, products, or functions; conducting joint annual planning sessions; and conducting joint risk or control sessions. Figure 2 summarizes hypothesized associations between the likelihood of a MW disclosure and variables measuring IAF attributes.

Control Variables

In Equation (1) we use the following variables to control for several firm-specific factors that prior research has shown to be correlated with the likelihood that a MW is disclosed: (1) *SEGMENTS* and *FOREIGNTRANSACTIONS* proxy for internal control risks that stem from organizational complexity; (2) *RESTRUCTURE*, *M&A*, and *SALESGROWTH* proxy for internal control risk associated with rapid organizational change; (3) *INVENTORY* captures internal control risk associated with inventory and product obsolescence costs; (4) *MARKETVALUE* controls for the level of resources available to invest in internal controls; (5) *LOSS*, *CFO*, and *SHUMWAY* measure financial distress and reflect a firm's diminished ability to adequately invest in internal controls;¹⁸

FIGURE 2
Hypothesized Relations between IAF Attributes and Activities and the Likelihood That a Firm Discloses a Material Weakness



¹⁸ We also used Altman's Z-score in place of the Shumway score to evaluate the sensitivity of the model. The signs and significance levels of IAF variables remain unchanged in our full model.

(6) *AGE* measures firm age; (7) *AUDITORSPECIALIST* is a dichotomous variable indicating whether the external auditor is an industry specialist; and (8) *BLUERIBBONAC* controls for the extent to which firms have implemented Blue Ribbon Committee recommendations (Carcello et al. 2005; Krishnan 2005; Ashbaugh-Skaife et al. 2007; Doyle et al. 2007a; Stephens 2009). We include *ISO9000* to control for firm certification by the International Standards Organization (ISO) because ISO implementation may potentially improve the internal controls associated with organizational business practices. The variables *RESTATEMENT* and *AUDITORRESIGN* control for existing internal control problems associated with earnings restatements and auditor resignations, respectively, during the 2001–2005 period. Finally, *INSTITUTIONALOWNERSHIP* and *REGULATEDINDUSTRY* proxy for external monitoring mechanisms that may be associated with the frequency of material weakness disclosures. Variable definitions are summarized in Table 2.

V. RESULTS

Descriptive Statistics

Table 3 provides descriptive statistics for the firms in our sample, along with comparative statistics for firms in the Compustat universe. Panel A of Table 3, which presents the distribution of sample firms by industry, indicates that the sample contains a high concentration of firms in regulated industries. Wilcoxon tests, reported in the extreme right column of Panel B, indicate that our sample firms are significantly larger, older, and more profitable than those in the Compustat universe. Prior research suggests that firms in regulated industries, and those that are older, larger, and more profitable, make greater investments in the IAF (Wallace and Kreutzfeldt 1991; Carcello et al. 2005; Goodwin and Kent 2006). The industry, size, and profitability comparisons are consistent with the propensity for GAIN survey respondents to represent firms that have relatively large internal audit functions. To the extent that firms with large internal audit functions are more prone to respond to the GAIN survey, our tests are potentially biased against detecting associations between IAF quality and material weaknesses.

Panel C of Table 3 compares the industry distribution of reported material weaknesses in our study sample to the distribution of firms reported in Doyle et al. (2007a). Approximately 21 percent of our sample firms disclosed *at least* one material weakness during the period November 2004 through December 2006. This is comparable to Doyle et al. (2007b), who find that 17 percent of their sample firms report material weaknesses during the period from August 2002 through October 2005.

Table 4 presents descriptive statistics for the independent variables in our model for the overall sample, partitioned on MW disclosures. Univariate tests of differences between the partitions indicate that material weakness firms are more likely to have internal auditors that issue grades in their audit reports (64 percent versus 42 percent; $p = 0.019$), and have fewer financial reporting activities audited by their internal auditors. Consistent with prior research, material weakness firms have significantly higher incidences of restructuring activities and losses, lower cash flows from operations, and higher bankruptcy risk. Univariate tests also show that a significantly greater proportion of material weakness firms report foreign currency adjustments and auditor resignations. Finally, the material weakness firms in our sample tend to be smaller and less likely to have *ISO9000* certifications. Overall, these results suggest that material weakness disclosures are influenced by differences in some IAF attributes in addition to previously documented firm characteristics.

Table 5 provides Spearman and Pearson correlation coefficients for the independent variables in our model. Since IAF measures attempt to capture the same underlying construct, quality, we find some significant correlations between IAF attributes and activities (Table 5, Panel A). In particular, two measures of competence, *CERTIFICATIONS* and *EXPERIENCE*, are highly correlated ($r = 0.39$). Also, *FINANCIALFOCUS* is significantly correlated with *CAEAC* ($r = 0.34$),

TABLE 3
Comparison of Sample Firms to Compustat Universe and Doyle et al. (2007a) Sample

Panel A: Distribution by Industry

Industry ^a	n	Sample (n = 214) %	2003 Compustat (n = 7,019) %
Banks and Insurance	38	17.76	20.80
Utilities	38	17.76	2.55
Durable Manufacturers	33	15.42	15.55
Computers	22	10.28	15.07
Textiles, Printing, and Publishing	20	9.35	3.36
Transportation	15	7.01	6.15
Retail	15	7.01	7.81
Services	8	3.74	8.32
Chemicals	9	4.21	2.14
Refining and Extractive	4	1.87	3.46
Food	6	2.80	1.97
Drug and Medical Equipment	4	1.87	9.06
Mining and Construction	1	0.47	2.45
Miscellaneous	1	0.47	1.01
Agriculture	—	—	0.30
Total	214	100.00	100.00

Panel B: Distribution by Firm Characteristics

	Sample (n = 214)		2003 Compustat (excluding sample firms) ^b		Median Comparison ^c
	Mean	Median	Mean	Median	
Market Value (millions)	14,542.16	4683.13	2,474.69	180.46	<0.001
Total Assets (millions)	25,601.42	7,835.65	6,208.13	226.90	<0.001
Firm Age	19.42	24.00	12.46	11.00	<0.001
Market-to-Book	2.85	2.04	2.92	1.94	0.013

(continued on next page)

Panel B: Distribution by Firm Characteristics

	Sample (n = 214)		2003 Compustat (excluding sample firms) ^b		Median Comparison ^c
	Mean	Median	Mean	Median	
ROA	0.036	0.032	-0.229	0.012	<0.001
Segments	3.53	3.00	2.86	2.00	<0.001

Panel C: Distribution of Sample Firms with Material Weaknesses by Industry

Industry ^d	Sample		Doyle et al. (2007a) Sample	
	n	%	n	%
Banks and Insurance	11	24.44	147	13.3
Durable Manufacturers	6	13.33	198	18.0
Utilities	6	13.33	35	3.2
Computers	5	11.11	208	18.9
Retail	4	8.89	137	12.4
Textiles, Printing, and Publishing	5	11.11	41	3.7
Transportation	3	6.67	64	5.8
Services	3	6.67	106	9.6
Drug and Medical Equipment	1	2.22	72	6.5
Food	1	2.22	16	1.4
Refining and Extractive	—	—	32	2.9
Mining and Construction	—	—	21	1.9
Chemicals	—	—	21	1.9
Miscellaneous	—	—	4	0.4
Total	45	100.00	1,102	100.0

^a Industry affiliations are determined using the following SIC codes: Agriculture: 100–999; Mining and Construction: 1000–1299, 1400–1999; Food: 2000–2111; Textiles, Printing, and Publishing: 2200–2799; Chemicals: 2800–2824, 2840–2899; Drug and Medical Equipment: 2830–2836, 3840–3851; Refining and Extractive: 1300–1399, 2900–2999; Durable Manufacturers: 3000–3569, 3580–3669, 3680–3839, 3852–3999; Computers: 3570–3579, 3670–3679, 7370–7379; Transportation: 4000–4899; Utilities: 4900–4999; Retail: 5000–5999; Banks and Insurance: 6000–6999; Services: 7000–7369, 7380–8999; and Miscellaneous: 9000–9999. The benchmark group consists of all publicly traded firms on 2003 Compustat with non-missing market value of equity (#199 times #25), total assets (#6), and earnings before extraordinary items (#18).

^b The benchmark group consists of a maximum of 6,803 firm observations from 2003 Compustat database after excluding our sample firms. Market Value is the market value of equity (#199 times #25); Total Assets (#6); Firm Age is the number of years the firm has Compustat data since 1980 and up to 2003; Market-to-Book is the market value of equity

(continued on next page)

(#199 times #25) divided by common equity (#60), winsorized at top and bottom 1 percent; ROA is earnings before extraordinary items (#18) divided by average total assets (#6), Winsorized at top and bottom 1 percent; Segments is the sum of the number of operating and geographic segments reported by the Compustat Segments database.

^c Reported p-values are from two-tailed Wilcoxon rank sum tests.

^d Industry affiliations are determined using the following SIC codes: Mining and Construction: 1000–1299, 1400–1999; Food: 2000–2111; Textiles, Printing, and Publishing: 2200–2799; Chemicals: 2800–2824, 2840–2899; Drug and Medical Equipment: 2830–2836, 3840–3851; Refining and Extractive: 1300–1399, 2900–2999; Durable Manufacturers: 3000–3569, 3580–3669, 3680–3839, 3852–3999; Computers: 3570–3579, 3670–3679, 7370–7379; Transportation: 4000–4899; Utilities: 4900–4999; Retail: 5000–5999; Banks and Insurance: 6000–6999; Services: 7000–7369, 7380–8999; and Miscellaneous (9000–9999). There are no material weakness firm observations in Agriculture (100–999) for both our sample and Doyle et al.'s (2007a) sample.

TABLE 4
Descriptive Statistics for IAF Attributes, IAF Activities, and Control Variables, Partitioned by Whether Firm Disclosed a Material Weakness^a

	Full Sample (n = 214)						
	Mean	Std. Dev.	Min	Q1	Median	Q3	Max
EXPERIENCE	8.901	4.343	1.000	6.000	8.000	11.000	24.000
EDUCATION	4.467	0.503	2.000	4.200	4.509	4.750	6.000
CERTIFICATIONS	0.625	0.236	0.000	0.500	0.613	0.800	1.000
TRAINING	53.640	28.600	0.000	40.000	48.000	74.000	178.000
CAEAC	0.822	0.181	0.000	0.750	0.875	1.000	1.000
CAEOFFICER	0.388	0.488	0.000	0.000	0.000	1.000	1.000
IASIZE	0.031	0.028	0.001	0.012	0.022	0.038	0.160
FIELDWORKQA	1.318	0.360	0.000	1.143	1.357	1.571	2.000
IAGRADE	0.467	0.500	0.000	0.000	0.000	1.000	1.000
FINANCIALFOCUS	1.185	0.447	0.000	0.833	1.167	1.500	2.000
FOLLOWUP	0.804	0.398	0.000	1.000	1.000	1.000	1.000
COORDINATION	0.883	0.322	0.000	1.000	1.000	1.000	1.000
SEGMENTS	3.528	2.764	1.000	1.000	3.000	5.000	19.000
FOREIGNTRANSACTIONS	0.299	0.459	0.000	0.000	0.000	1.000	1.000
M&A	0.193	0.243	0.000	0.000	0.200	0.400	1.000
RESTRUCTURE	0.379	0.367	0.000	0.000	0.400	0.600	1.000
SALESGROWTH	0.105	0.166	0.000	0.000	0.000	0.200	1.000
INVENTORY	0.075	0.095	0.000	0.009	0.041	0.108	0.602
MARKETVALUE (\$M)	14,542	31,946	90	1,763	4,683	11,610	276,168
LOSS	0.121	0.231	0.000	0.000	0.000	0.200	1.000
CFO	0.088	0.057	-0.100	0.050	0.081	0.122	0.254
SHUMWAY	0.004	0.010	0.000	0.001	0.002	0.004	0.139
AGE	19.421	6.440	4.000	14.000	24.000	24.000	24.000
AUDITORSPECIALIST	0.383	0.487	0.000	0.000	0.000	1.000	1.000
BLUERIBBONAC	0.973	0.093	0.100	1.000	1.000	1.000	1.000
ISO9000	0.416	0.494	0.000	0.000	0.000	1.000	1.000

(continued on next page)

Full Sample
(n = 214)

	Mean	Std. Dev.	Min	Q1	Median	Q3	Max
RESTATEMENT	0.706	0.457	0.000	0.000	1.000	1.000	1.000
AUDITORRESIGN	0.042	0.201	0.000	0.000	0.000	0.000	1.000
INSTITUTIONALOWNERSHIP	0.461	0.135	0.098	0.370	0.465	0.547	0.917
REGULATEDINDUSTRY	0.355	0.480	0.000	0.000	0.000	1.000	1.000

Non-MW Firms
(n = 169)

Material Weakness Firms (n = 45)

	Material Weakness Firms (n = 45)		Non-MW Firms (n = 169)		p-values for Mean/Median Comparisons ^b	
	Mean	Median	Mean	Median	Mean	Median
EXPERIENCE	9.420	9.000	8.763	8.000	0.369	0.285
EDUCATION	4.379	4.400	4.491	4.542	0.187	0.226
CERTIFICATIONS	0.652	0.615	0.618	0.600	0.397	0.436
TRAINING	52.311	45.000	53.994	50.000	0.767	0.409
CAEAC	0.786	0.875	0.831	0.875	0.136	0.193
CAEOFFICER	0.422	0.000	0.379	0.000	0.596	0.596
IASIZE	0.030	0.021	0.032	0.022	0.693	0.944
FIELDWORKQA	1.244	1.143	1.338	1.429	0.196	0.145
IAGRADE	0.622	1.000	0.426	0.000	0.019	0.019
FINANCIALFOCUS	1.067	1.000	1.216	1.167	0.046	0.063
FOLLOWUP	0.756	1.000	0.817	1.000	0.362	0.362
COORDINATION	0.933	1.000	0.870	1.000	0.168	0.241
SEGMENTS	3.222	2.000	3.609	3.000	0.405	0.228
FOREIGNTRANSACTIONS	0.400	0.000	0.272	0.000	0.097	0.097
M&A	0.244	0.200	0.179	0.200	0.108	0.229
RESTRUCTURE	0.476	0.400	0.354	0.200	0.048	0.082
SALESGROWTH	0.111	0.000	0.103	0.000	0.776	0.558
INVENTORY	0.076	0.022	0.075	0.043	0.932	0.330
MARKETVALUE (\$M)	7,158	2,863	16,508	5,089	0.005	0.021
LOSS	0.227	0.000	0.093	0.000	0.008	0.003
CFO	0.060	0.056	0.095	0.085	<0.0001	0.000

(continued on next page)

	Material Weakness Firms (n = 45)		Non-MW Firms (n = 169)		p-values for Mean/Median Comparisons ^b	
	Mean	Median	Mean	Median	Mean	Median
SHUMWAY	0.008	0.003	0.003	0.002	0.082	0.057
AGE	19.444	24.000	19.414	24.000	0.978	0.935
AUDITORSPECIALIST	0.356	0.000	0.391	0.000	0.670	0.670
BLUERIBBONAC	0.971	1.000	0.973	1.000	0.885	0.704
ISO9000	0.289	0.000	0.450	0.000	0.052	0.053
RESTATEMENT	0.667	1.000	0.716	1.000	0.521	0.521
AUDITORRESIGN	0.111	0.000	0.024	0.000	0.079	0.010
INSTITUTIONALOWNERSHIP	0.456	0.470	0.463	0.460	0.749	0.955
REGULATEDINDUSTRY	0.378	0.000	0.349	0.000	0.723	0.723

^a See Table 2 for variable definitions.

^b Reported p-values for means are from two-tailed t-tests; reported p-values for medians are from two-tailed Wilcoxon rank sum tests.

TABLE 5
Spearman/Pearson Correlations Among Measures of IAF Quality and Control Variables (n = 214)

Panel A: Spearman/Pearson Correlations among Measures of IAF Quality									
MW	EXPERIENCE	EDUCATION	CERTIFICATION	TRAINING	CAEAC				
EXPERIENCE	0.07	-0.09	0.06	-0.02	-0.10				
EDUCATION	-0.08	-0.11	0.38	-0.09	0.01				
CERTIFICATIONS	0.05		-0.02	0.15	-0.04				
TRAINING	-0.06	-0.01		-0.14	0.02				
CAEAC	-0.09	0.22	-0.10	0.12	0.13				
CAEOFFICER	0.04	0.00	0.04	0.13	0.12				
IASIZE	-0.01	-0.06	-0.07	-0.14	0.19				
FIELDWORKQA	-0.10	0.13	-0.08	0.18	0.13				
IAGRADE	-0.10	0.00	-0.05	0.06	0.35				
FINANCIALFOCUS	0.16	0.01	-0.11	0.16	0.12				
FOLLOWUP	-0.13	0.02	0.01	0.04	0.14				
COORDINATION	-0.06	-0.01	0.04	0.17	0.13				
	0.08	-0.10	-0.03						
Panel B: Spearman/Pearson Correlations among Measures of IAF Quality									
MW	EXPERIENCE	EDUCATION	CERTIFICATION	TRAINING	CAEAC				
EXPERIENCE	0.04	-0.03	-0.14	-0.06	0.08				
EDUCATION	-0.06	0.05	0.01	0.01	-0.13				
CERTIFICATIONS	-0.06	0.15	0.00	-0.03	-0.02				
TRAINING	-0.07	0.00	0.00	0.05	-0.04				
CAEAC	0.09	-0.14	0.16	0.04	0.17				
CAEOFFICER	0.12	0.12	0.34	0.13	0.20				
IASIZE	-0.22	-0.21	-0.04	0.08	0.08				
FIELDWORKQA	-0.10	-0.10	0.01	0.01	-0.10				
IAGRADE	-0.04	0.06	0.26	0.13	0.23				
FINANCIALFOCUS	0.08	0.12	0.12	0.09	0.02				
FOLLOWUP	0.06	0.05	0.02	0.02	0.15				
COORDINATION	-0.11	-0.10	0.12	0.15					

(continued on next page)

Panel B: Spearman Correlations among Measures of IAF Quality and Control Variables

	MW	EXPERIENCE	EDUCATION	CERTIFICATION	TRAINING	CAEAC
SEGMENTS	-0.08	-0.07	0.15	-0.03	-0.07	0.02
FOREIGNTRANSACTIONS	0.11	-0.20	0.11	-0.11	-0.13	-0.02
M&A	0.08	-0.03	0.02	-0.03	0.06	0.06
RESTRUCTURE	0.12	-0.18	0.07	-0.13	0.02	0.08
SALESGROWTH	0.04	0.16	-0.02	0.06	0.04	0.02
INVENTORY	-0.07	-0.14	0.01	-0.11	0.03	-0.05
MARKETVALUE	-0.16	-0.22	0.17	-0.19	0.31	0.15
LOSS	0.20	-0.11	0.04	0.02	-0.04	-0.06
CFO	-0.26	-0.13	0.12	-0.01	-0.10	0.04
SHUMWAY	0.13	0.04	0.00	-0.05	0.17	0.00
AGE	-0.01	-0.05	0.04	-0.09	0.16	-0.09
AUDITORSPECIAL	-0.03	0.02	0.05	-0.01	0.13	0.08
BLUERIBBONAC	-0.03	-0.03	-0.03	-0.07	0.17	0.17
ISO9000	-0.13	-0.27	0.09	-0.18	0.11	0.02
RESTATEMENT	-0.04	-0.02	0.03	0.04	0.08	0.07
AUDITORRESIGN	0.18	0.02	-0.06	0.04	-0.04	0.08
INSTITUTIONOWNER	0.00	0.06	0.01	0.03	-0.14	-0.03
REGULATEDINDUSTRY	0.02	0.28	-0.09	0.11	0.14	0.02

	CAEOFFICER	IASIZE	FIELDWORKQA	IAGRADE	FINANCIALFOCUS	FOLLOWUP	COORDINATION
SEGMENTS	0.19	-0.01	0.01	0.02	0.01	0.00	0.02
FOREIGNTRANSACTIONS	0.20	-0.02	0.08	0.10	-0.03	-0.04	-0.02
M&A	0.14	-0.03	-0.07	0.08	-0.02	0.00	-0.04
RESTRUCTURE	0.18	-0.04	0.06	0.16	-0.10	0.01	0.05
SALESGROWTH	0.02	0.07	-0.15	0.02	0.12	-0.02	-0.10
INVENTORY	0.42	0.10	-0.07	0.05	-0.09	0.06	-0.07
MARKETVALUE	-0.28	-0.46	0.26	0.16	0.05	0.06	0.18
LOSS	-0.01	-0.10	0.09	0.11	-0.03	-0.02	-0.02
CFO	0.40	0.09	-0.05	-0.18	0.02	-0.06	-0.10
SHUMWAY	-0.20	-0.05	0.15	0.16	0.06	0.06	0.15
AGE	0.03	-0.15	0.14	-0.02	-0.07	0.05	0.02
AUDITORSPECIAL	-0.09	-0.09	0.10	0.01	-0.04	0.17	0.14

(continued on next page)

	CAEOFFICER	IASIZE	FIELDWORKQA	IAGRADE	FINANCIALFOCUS	FOLLOWUP	COORDINATION
BLUERIBBONAC	-0.08	-0.07	0.24	0.04	0.10	-0.02	0.20
ISO9000	0.29	0.01	0.11	0.07	-0.02	0.04	-0.05
RESTATEMENT	-0.18	-0.13	0.08	0.01	-0.05	0.04	-0.04
AUDITORRESIGN	0.06	0.09	0.02	0.04	0.07	0.05	0.00
INSTITUTIONOWNER	0.20	0.05	-0.05	-0.13	-0.09	0.01	0.05
REGULATEDINDUSTRY	-0.50	-0.02	-0.01	0.03	0.03	0.05	0.06

Panel C: Pearson Correlations among Measures of IAF Quality and Control Variables

	MW	EXPERIENCE	EDUCATION	CERTIFICATION	TRAINING	CAEAC
SEGMENTS	-0.08	-0.09	0.12	-0.05	-0.08	0.01
FOREIGNTRANSACTIONS	0.11	-0.16	0.12	-0.09	-0.09	-0.04
M&A	0.11	-0.05	0.04	-0.05	0.09	0.07
RESTRUCTURE	0.14	-0.16	0.07	-0.11	0.02	0.04
SALES&GROWTH	0.02	0.14	0.01	0.05	0.02	0.02
INVENTORY	0.01	-0.05	-0.06	-0.03	-0.03	0.00
MARKETVALUE	-0.17	-0.21	0.16	-0.17	0.26	0.18
LOSS	0.24	-0.08	0.00	0.09	-0.06	-0.04
CFO	-0.25	-0.08	0.11	-0.02	-0.13	0.04
SHUMWAY	0.23	-0.03	-0.01	0.01	-0.10	-0.02
AGE	0.01	-0.05	-0.02	-0.07	0.12	-0.11
AUDITOR&SPECIAL	-0.03	0.01	0.06	-0.01	0.10	0.08
BLUERIBBONAC	-0.01	-0.04	-0.03	-0.01	0.10	0.31
ISO9000	-0.13	-0.20	0.04	-0.18	0.11	0.00
RESTATEMENT	-0.04	-0.06	0.01	0.06	0.11	0.07
AUDITORRESIGN	0.18	0.01	0.01	0.04	-0.04	0.08
INSTITUTIONOWNER	-0.02	0.06	0.08	0.07	-0.07	-0.02
REGULATEDINDUSTRY	0.02	0.19	-0.10	0.11	0.11	0.04

	CAEOFFICER	IASIZE	FIELDWORKQA	IAGRADE	FINANCIALFOCUS	FOLLOWUP	COORDINATION
SEGMENTS	-0.09	0.10	0.01	0.03	0.01	0.00	0.02
FOREIGNTRANSACTIONS	-0.12	0.11	0.07	0.10	-0.03	-0.04	-0.02
M&A	-0.05	0.03	0.00	0.06	0.01	-0.01	-0.05

(continued on next page)

	CAEOFFICER	IASIZE	FIELDWORKQA	IAGRADE	FINANCIALFOCUS	FOLLOWUP	COORDINATION
RESTRUCTURE	-0.01	0.06	0.04	0.19	-0.10	0.00	0.04
SALESGROWTH	0.03	-0.06	-0.11	0.02	0.09	0.01	-0.05
INVENTORY	-0.10	0.35	-0.14	-0.02	-0.11	-0.02	-0.07
MARKETVALUE	0.23	-0.35	0.25	0.15	0.05	0.07	0.17
LOSS	-0.14	0.01	0.07	0.14	-0.03	-0.03	-0.09
CFO	-0.20	0.36	-0.04	-0.18	0.04	-0.03	-0.11
SHUMWAY	0.01	-0.09	0.04	<i>0.13</i>	-0.10	-0.09	0.00
AGE	-0.13	-0.14	<i>0.12</i>	-0.06	-0.08	0.03	0.05
AUDITORSPECIAL.	0.06	-0.14	0.09	0.01	-0.04	0.17	0.14
BLUERIBBONAC	0.02	-0.03	0.25	0.10	0.03	0.11	0.22
ISO9000	-0.09	0.14	0.09	0.07	0.00	0.04	-0.05
RESTATEMENT	0.05	-0.20	0.06	0.01	-0.03	0.04	-0.04
AUDITORRESIGN	-0.07	0.09	0.00	0.04	0.08	0.05	0.00
INSTITUTIONOWNER	-0.13	0.15	-0.07	-0.12	-0.12	-0.01	0.05
REGULATEDINDUSTRY	0.21	-0.41	0.03	0.03	0.03	0.05	0.06

Bold = p-value < 0.05. Italic = p-value < 0.10.

See Table 2 for variable definitions.

suggesting that IAFs that focus greater attention on financial reporting activities tend to report more information to the audit committee. All other significant correlations among our IAF quality measures fall below 0.30. One measure of objectivity, *CAEOFFICER*, is significantly associated with three control variables, *INVENTORY* (Spearman $r = 0.42$), *CFO* (Spearman $r = 0.40$), and *REGULATEDINDUSTRY* (Spearman $r = -0.50$). Investment in the IAF (*IASIZE*) is significantly and negatively correlated with the two control variables, *MARKETVALUE* (Spearman $r = -0.46$; Pearson $r = -0.35$) and *REGULATEDINDUSTRY* (Pearson $r = -0.41$). Remaining correlations between IAF quality measures and control variables are below 0.40.

Multivariate Results

Table 6 reports the results of our logistic regression. We report parameter estimates, marginal effects, and the change in the probability of a firm disclosing a material weakness as a result of moving from the first to the third quartile value of the independent variable, while holding other regressors at their mean values.¹⁹ We report the results for full and stepwise regression models in Panels A and B, respectively. All p-values refer to two-tailed tests of significance.

We find little support for H1, which predicts associations between MW disclosures and the IAF quality attributes of competence, objectivity, and IAF investment. Among our four proxies for IAF competence, only *EDUCATION* is significantly associated with the probability that a firm reports a material weakness (Coeff. = -1.682 , $p < 0.05$). The results in column (4) indicate that moving from first to third quartile of *EDUCATION* decreases the probability of material weakness disclosure by 1.8 percent. Contrary to H1, the remaining measures of IAF competence (*EXPERIENCE*, *CERTIFICATION*, and *TRAINING*) and our proxies for IAF objectivity (*CAEAC* or *CAEOFFICER*) and IAF investment (*IASIZE*) are not significantly associated with MW disclosures. The lack of a statistically significant association could reflect three factors. First, IAF competence, objectivity, and investment improve the IAF's ability to prevent material weaknesses from occurring, while also increasing the likelihood that existing material weaknesses are detected and disclosed. Hence, these opposing effects may offset each other. Second, the small size and relative homogeneity of our sample, combined with the large number of control variables, lowers the power of our statistical tests. Finally, *CAEAC* and *IASIZE* are significantly correlated with multiple control variables that are significant predictors of MWs, making it difficult to isolate relations between these IAF attributes and MW disclosures. Accordingly, further investigation into these specific aspects of IAF quality is warranted.

We find strong support for H2, which predicts associations between MW disclosures and the following IAF activities: *FIELDWORKQA*, *FINANCIALFOCUS*, and *IAGRADE*. The coefficients on *FIELDWORKQA* and *FINANCIALFOCUS* are negative and significant (Coeff. = -2.674 , $p < 0.01$; Coeff. = -1.955 , $p < 0.05$, respectively), indicating that MW disclosures are decreasing in the use of quality assurance techniques during fieldwork and the extent to which IAF scope includes financial reporting processes. Moving from first to third quartile of *FIELDWORKQA* and *FINANCIALFOCUS* decreases the probability of firms reporting a material weakness by 2.1 percent and 2.7 percent, respectively. We also find that *IAGRADE* is positively and significantly associated with the probability of disclosing a material weakness (Coeff. = 1.445 , $p < 0.05$), suggesting that grades facilitate Section 404 risk assessment, and thereby increase the likelihood that extant MWs are identified. Firms whose internal auditors issue grades or ratings are 3.1 percent more likely to report a material weakness than those that do not.

¹⁹ We compute the marginal effects as $e^{\beta'X} / (1 + e^{\beta'X})^2$ where $\beta'X$ is computed using the mean value of X . If X contains logs, then we substitute the mean of the original variable into the logarithm function.

TABLE 6
Logistic Model Estimating the Relation between IAF Attributes and Activities and the Likelihood of a
Material Weakness Disclosure^a

Panel A: Full Model

	Coeff. (χ^2) (1)	Marginal Effect ^c (2)	Change in Prob. Q1 versus Q3 Values ^d (3)
Intercept	6.780 (2.36)		
IAF Attributes			
<i>EXPERIENCE</i>	0.001 (0.00)	0.000	0.000
<i>EDUCATION</i>	-1.682** (6.19)	-0.032	-0.018
<i>CERTIFICATIONS</i>	0.933 (0.53)	0.018	0.005
<i>TRAINING</i>	0.001 (0.01)	0.000	0.001
<i>CAEAC</i>	-1.660 (0.97)	-0.031	-0.007
<i>CAEOFFICER</i>	0.412 (0.47)	0.008	0.008
<i>IASIZE</i>	6.523 (0.18)	0.124	0.003
IAF Activities			
<i>FIELDWORKQA</i>	-2.674*** (7.93)	-0.051	-0.021
<i>IAGRADE</i>	1.445** (5.64)	0.027	0.031
<i>FINANCIALFOCUS</i>	-1.955** (6.12)	-0.037	-0.027
<i>FOLLOWUP</i>	-1.213* (0.18)	-0.023	-0.034

(continued on next page)

Panel A: Full Model

	Coeff. (χ^2) (1)	Marginal Effect ^c (2)	Change in Prob. Q1 versus Q3 Values ^a (3)
<i>COORDINATION</i>	(3.63) 4.095*** (8.00)	0.078	0.030
Control Variables			
<i>SEGMENTS</i>	-1.465*** (8.95)	-0.028	-0.099
<i>FOREIGNTRANSACTIONS</i>	3.650*** (15.13)	0.069	0.196
<i>M&A</i>	4.629*** (11.60)	0.088	0.041
<i>RESTRUCTURE</i>	1.137 (1.37)	0.022	0.012
<i>SALESGROWTH</i>	-1.093 (0.42)	-0.021	-0.004
<i>INVENTORY</i>	-7.645** (3.88)	-0.145	-0.017
<i>MARKETVALUE</i>	-0.387 (2.15)	-0.007	-0.022
<i>LOSS</i>	1.793 (1.33)	0.034	0.007
<i>CFO</i>	-28.061*** (9.87)	-0.532	-0.046
<i>SHUNWAY</i>	210.300** (4.11)	3.987	0.008
<i>AGE</i>	2.327*** (7.43)	0.044	0.022
<i>AUDITORSPECIALIST</i>	0.863 (1.89)	0.016	0.019
<i>BLUERIBBONAC</i>	0.529	0.010	0.000

(continued on next page)

Panel A: Full Model

	Coeff. (χ^2) (1)	Marginal Effect ^c (2)	Change in Prob. Q1 versus Q3 Values ^d (3)
ISO9000	(0.04) -4.006*** (14.84)	-0.076	-0.093
RESTATEMENT	-0.401 (0.45)	-0.008	-0.008
AUDITORRESIGN	5.492*** (12.95)	0.104	0.776
INSTITUTIONALOWNERSHIP	-1.022 (0.19)	-0.019	-0.003
REGULATEDINDUSTRY	-1.723* (2.98)	-0.033	-0.029
Likelihood Ratio χ^2	107.15		
Model Significance	<0.0001		
Max-Rescaled R ² (%)	61.31		
Number of MW Obs.	45		
Total Number of Obs.	214		

Panel B: Stepwise Model^b

	Coeff. (χ^2) (4)	Marginal Effect ^c (5)	Change in Prob. Q1 versus Q3 Values ^d (6)
Intercept	7.720 (5.09)		
IAF Attributes EDUCATION	-1.579*** (7.19)	-0.053	-0.029

(continued on next page)

Panel B: Stepwise Model^b

	Coeff. (χ^2) (4)	Marginal Effect ^c (5)	Change in Prob. Q1 versus Q3 Values ^d (6)
IAF Activities			
FIELDWORKQA	-2.324*** (9.02)	-0.078	-0.032
IAGRADE	1.500*** (6.81)	0.050	0.056
FINANCIALFOCUS	-2.107*** (9.94)	-0.070	-0.052
FOLLOWUP	-1.137* (3.67)	-0.038	-0.054
COORDINATION	3.163*** (7.72)	0.106	0.047
Control Variables			
SEGMENTS	-1.376*** (9.66)	-0.046	-0.147
FOREIGNTRANSACTIONS	3.642*** (16.30)	0.122	0.303
M&A	4.311*** (12.03)	0.144	0.065
RESTRUCTURE	1.196 (1.90)	0.040	0.022
INVENTORY	-7.720** (5.33)	-0.258	-0.029
MARKETVALUE	-0.487** (4.92)	-0.016	-0.031
CFO	-29.530*** (13.65)	-0.986	-0.085
SHUMWAY	238.700** (7.32)	7.974	0.016

(continued on next page)



Panel B: Stepwise Model^b

	Coeff. (χ^2) (4)	Marginal Effect ^c (5)	Change in Prob. Q1 versus Q3 Values ^d (6)
AGE	2.226*** (8.18)	0.074	0.037
AUDITORSPECIALIST	0.721 (1.71)	0.024	0.026
ISO9000	-3.928*** (16.90)	-0.131	-0.152
AUDITORRESIGN	5.102*** (13.73)	0.170	0.798
REGULATEDINDUSTRY	-2.183** (6.39)	-0.073	-0.064
Likelihood Ratio χ^2 Model	102.88		
Significance	<0.0001		
Max-Rescaled χ^2 (%)	59.41		
Number of MW Obs.	45		
Total Number of Obs.	214		

***, **, * Pr > χ^2 of 0.1, 0.05, and 0.01, respectively.

^a See Table 2 for variable definitions.

^b The significance level of χ^2 for a variable to stay in the model is 0.30.

^c The marginal effects are computed as $e^{\beta'X} / (1 + e^{\beta'X})^2$, where $\beta'X$ is computed using the mean values of X. If X contains logs, then we substitute means of original variables into log functions.

^d Tabled values in the Change in Probability column show the change in the probability of a firm disclosing a material weakness as a result of moving from the first to the third quartile value of the variable of interest or from 0 to 1 if the variable is an indicator variable, holding all other variables constant at their mean values.

As predicted by H3, firms with IAFs that follow-up on extant control problems are significantly (Coeff. = -1.213 , $p < 0.10$) less likely to report a material weakness. This result suggests that monitoring the remediation of previously identified control problems prompts management to correct them in a timely manner. A one-unit change in *FOLLOWUP* is associated with a 3.4 percent decrease in the probability of reporting a material weakness.

Consistent with H4, firms are significantly more likely to report a material weakness (Coeff. = 4.095 , $p < 0.01$) when the IAF coordinates audit activities with the external auditors. This result suggests that external auditors can increase the effectiveness of their Section 404 procedures by coordinating with a firm's IAF. The change in probability analysis shows that a unit change in *COORDINATION* increases the probability of firms reporting a material weakness by 3 percent.

Our findings regarding the nature and scope of IAF activities suggest implications for practice. The results for *FIELDWORKQA* and *FINANCIALFOCUS* indicate that these audit practices prevent MWs from occurring and/or ensure that extant MWs are detected early enough in the firm's Section 404 testing to enable management to correct problems. The year-end timing of most Section 404 work suggests that these activities are more preventative than detective. Together with tests of the *FOLLOWUP* variable, our results imply that firms can potentially improve ICFR by expanding fieldwork quality assurance practices, devoting greater effort to financial statement reporting processes, and aggressively following up on previously detected audit exceptions.

The positive relations between MW disclosures and both *IAGRADE* and *COORDINATION* suggest that these activities increase the effectiveness of Section 404 compliance processes by facilitating risk assessment. IAF grading promotes rapid assessment and prioritization of control risks by management, audit committees, and external auditors (PwC 2006). Similarly, information gained through coordination enables external auditors to alter the nature, timing, and extent of their testing to reflect risk levels. Accordingly, these findings confirm the appropriateness of recent regulatory guidance that instructs firms and external auditors to use a risk-based approach to Section 404 compliance (SEC 2007; PCAOB 2005, 2007b). The results are also consistent with recent PCAOB (2005, 2007a) guidance that encourages external auditors to more effectively use the work of others.²⁰

The signs and significance levels of our control variables are generally consistent with prior research, except for the negative coefficient on *SEGMENTS* and the positive coefficient on *AGE* (Ashbaugh-Skaife et al. 2007; Doyle et al. 2007a; Krishnan and Vishwanathan 2007; Stephens 2009). We attribute these differences to our sample, which has a much larger proportion of utilities and financial firms than samples used in prior studies.

Since many independent variables are not statistically significant in Panel A of Table 6 and we have a relative small sample size, we also present logit results for a stepwise hierarchical regression model in Panel B.²¹ The coefficient signs in the reduced model are consistent with the full model, while the sizes, significance levels, and marginal effects of the coefficients are generally larger.

Sensitivity Analysis

One limitation of our study is that the IAF data describe different years for different firms. IAF data for 57 firms in the sample correspond to the pre-SOX time period. This could bias our

²⁰ The methods of coordination used (multiple methods per firm are permitted) and the percentage of firms using each method are as follows: (1) loan to external auditors, 28 percent, (2) perform complete or partial audit of specific locations, products or functions, 58 percent, (3) conduct joint annual planning sessions, 58 percent, and (4) conduct joint risk or control sessions, 39 percent. We classify IAFs that responded "yes" to (1) and/or (2) as performing work for use by external auditors. We perform robustness tests with a continuous measure based on the number of coordination methods used. This variable gives similar results (Coeff. = 2.179 , $p = 0.046$). However, multiple methods of coordination are not necessarily indicative of more coordination; thus, we use an indicator variable in our main results.

²¹ We used a p-value of 0.30 as a discretionary cutoff for a variable to be included in the model. Additional sensitivity tests using cut-off p-values of 0.2, 0.4, and 0.5 produce materially similar results.

results if the year of IAF data collection is correlated with both MW disclosures and our measures of IAF quality. To test for bias, we re-estimate Equation (1) after including an indicator variable, *SOX*, to indicate whether IAF data were collected on or after July 2002. The coefficient on *SOX* (Coeff. = -0.418) is not statistically significant ($p = 0.509$), and inclusion of the *SOX* variable does not materially affect our results. For firms with survey data available for both 2003 and 2004, we use the data from 2003 (rather than 2004) and re-estimate the models. The results are largely consistent with our reported results.

We further confirm the robustness of our results by using additional control variables for external monitoring. We include an additional corporate governance factor, Gompers et al.'s (2003) *GINDEX*. *GINDEX* is measured as the average G-index score for 2002 and 2004. This reduces the sample size to 200 firms. We obtain similar results for all IAF variables except *IAGRADE* and *FOLLOWUP*. In the full model, the coefficients for *IAGRADE* and *FOLLOWUP* are insignificant. We also control for the possibility that a firm audited by one of the Big 4 audit firms may be more likely to disclose a material weakness. Our results remain unchanged from those reported in Table 6.

VI. CONCLUSION

This study investigates associations between material weakness disclosures and various IAF attributes and activities using survey data collected by the IIA. Our results indicate that the nature and scope of IAF activities are more strongly associated with MW disclosures than the IAF attributes of competence, objectivity, and investment. Among IAF attribute measures, only the education level of the IAF is significantly associated with MW disclosures. Regarding IAF activities, we find that MW disclosures are negatively associated with the extent to which the IAF uses QA techniques in fieldwork, audits activities related to financial reporting, and follows up on previously identified control problems. The year-end timing of most Section 404 work and the nature of follow-up procedures suggests that these activities are more likely to be preventative rather than detective. We also find that MW disclosures are positively related with both IAF grading of audit engagements and external-internal auditor coordination. We interpret this finding to indicate that these activities increase the effectiveness of Section 404 compliance processes by facilitating risk assessment, consistent with the risk-based approach promoted by regulatory guidance (SEC 2007; PCAOB 2005, 2007a). Together, our results have important implications for managers who determine IAF staffing and activities, standard-setters who provide auditing guidance, and external auditors responsible for Section 404 work.

This study makes several important contributions to the literature. First, we expand extant research on both internal auditing and ICFR by using survey data from companies to document associations between various measures of IAF quality and material weakness disclosures. Research on the determinants of internal control deficiencies has investigated firm characteristics such as size, profitability, and complexity, without consideration of the role of the IAF (Ashbaugh-Skaife et al. 2007; Doyle et al. 2007a). With the exception of Prawitt et al. (2009), prior IAF studies focus on how measures of IAF quality affect the quality of internal auditors' decisions, without linking IAF quality to an actual outcome measure of control effectiveness, such as a material weakness (Berry et al. 1987; Harrell et al. 1989; Church and Schneider 1995). Accordingly, this study complements Prawitt et al. (2009), which finds a positive relation between a comprehensive measure of IAF quality and earnings quality. In combination, these studies suggest that the IAF is an important component of the financial reporting process. Second, the bi-directional relations between IAF quality measures and MW disclosures provide evidence on the roles that specific IAF attributes and activities play in the existence and detection of MWs. Few archival studies directly link auditor practices to the prevention or detection of audit exceptions. Finally, we provide evidence that external auditors are more likely to detect material weaknesses

when they coordinate their efforts with the IAF. While a significant branch of the internal auditing literature focuses on the relation between external auditors and the IAF, no study of which we are aware has examined whether IAF involvement in external audits increases the effectiveness of external audits.

Our study is subject to several limitations. Most notably, the small size and homogeneity of our sample, combined with the large number of control variables, lowers the power of our statistical tests. Accordingly, we cannot determine whether the lack of statistically significant hypothesized relations between IAF attributes is due to low statistical power or competing effects on the existence and detection of control problems. The stepwise tests mitigate this problem but do not eliminate it. Furthermore, large firms with relatively sophisticated IAFs tend to participate in the GAIN survey. This limits our ability to generalize findings to firms that did not respond to the GAIN survey. Despite these limitations, this study increases our understanding of the IAF's role in ICFR.

REFERENCES

- Albrecht, W. S., and C. O. Albrecht. 2004. *Fraud Examination and Prevention*. Mason, Ohio: Thompson/South-Western.
- American Institute of Certified Public Accountants (AICPA). 1991. *The Auditor's Consideration of the Internal Audit Function in an Audit of Financial Statements*. Statement on Auditing Standards No. 65. New York, NY: AICPA.
- Ashbaugh-Skaife, H., D. Collins, and W. Kinney. 2007. The discovery and reporting of internal control deficiencies prior to SOX-mandated audits. *Journal of Accounting and Economics* 44 (1–2): 166–192.
- Asare, S. K., L. A. Cunningham, and A. Wright. 2007. The Sarbanes-Oxley Act: Legal implications and research opportunities. Research Paper 122, Boston College Law School Legal Studies Research Paper Series. Boston, MA: Boston College.
- Bailey, A. D., Jr., A. A. Gramling, and S. Ramamoorti, eds. 2003. *Research Opportunities in Internal Auditing*. Altamonte Springs, FL: IIA Research Foundation.
- Bedard, J. C., and L. Graham. 2011. Factors affecting the severity of Sarbanes-Oxley Section 404 internal control deficiencies: Archival evidence. *The Accounting Review* 86 (3): forthcoming.
- , R. Hoitash, and U. Hoitash. 2009. Evidence from the United States on the effect of auditor involvement in assessing internal control over financial reporting. *International Journal of Auditing* 13 (2): 105–125.
- Berry, L. E., G. B. Harwood, and J. L. Katz. 1987. Performance of auditing procedures by governmental auditors: Some preliminary evidence. *The Accounting Review* 58 (1): 14–28.
- Brown, P. R. 1983. Independent auditor judgment in the evaluation of internal audit functions. *Journal of Accounting Research* 21 (2): 444–455.
- , and V. Karan. 1986. One approach for assessing the operational nature of auditing standards: An analysis of SAS 9. *Auditing: A Journal of Practice & Theory* 6 (1): 134–147.
- Brown, L., and A. S. Pinello. 2007. To what extent does the financial reporting process curb earnings surprise games? *Journal of Accounting Research* 45 (5): 947–981.
- Carcello, J., D. R. Hermanson, and K. Raghunandan. 2005. Factors associated with U.S. public companies investment in internal auditing. *Accounting Horizons* 19 (2): 69–84.
- Clark, M. E., T. E. Gibbs, and R. G. Schroeder. 1980. Evaluating internal audit departments under SAS No. 9: Criteria for judging competence, objectivity, and performance. *The Woman CPA*. 22 (July): 8–11.
- Church, B. K., and A. Schneider. 1995. Internal auditors' memory for financial-statement errors. *Behavioral Research in Accounting* 7: 17–36.
- Davidson, R., J. Goodwin-Stewart, and P. Kent. 2005. Internal governance structures and earnings management. *Accounting and Finance* 45 (2): 241–267.
- Doyle, J., W. Ge, and S. McVay. 2007a. Determinants of weaknesses in internal control over financial reporting. *Journal of Accounting and Economics* 44 (1–2): 193–223.

- , ———, and ———. 2007b. Accruals quality and internal control over financial reporting. *The Accounting Review* 82 (5): 1141–1170.
- Ettredge, M. L., C. Li, and L. Sun. 2006. The Impact of SOX Section 404 Internal Control Quality Assessment on Audit Delay in the SOX Era. *Auditing: A Journal of Practice & Theory* 25 (2): 1–23.
- Frankel, R. M., M. F. Johnson, and K. K. Nelson. 2002. The relation between auditors' fees for nonaudit services and earnings management. *The Accounting Review* 77 (s-1): 71–105.
- Ge, W., and S. McVay. 2005. The disclosure of material weaknesses in internal control after the Sarbanes-Oxley Act. *Accounting Horizons* 19: 137–158.
- Gompers, P., J. Ishii, and A. Metrick. 2003. Corporate governance and equity prices. *The Quarterly Journal of Economics* 118 (1): 107–155.
- Goodwin, J., and P. Kent. 2006. Factors affecting the voluntary use of internal audit. *Managerial Auditing Journal* 21 (1): 81–101.
- Gramling, A. A., M. J. Maletta, A. Schneider, and B. K. Church. 2004. The role of the internal audit function in corporate governance: A synthesis of the extant internal auditing literature and directions for future research. *Journal of Accounting Literature* 23: 194–244.
- Harrell, A., M. Taylor, and E. Chewning. 1989. An examination of management's ability to bias the professional objectivity of internal auditors. *Accounting, Organizations and Society* 14 (3): 259–269.
- Hogan, C. E., and M. S. Wilkins. 2008. Evidence on the audit risk model: Do auditors increase audit fees in the presence of internal control deficiencies?. *Contemporary Accounting Research* 25 (1): 219–242.
- Hoitash, U., R. Hoitash, and J. C. Bedard. 2009. Corporate governance and internal control over financial reporting: A comparison of regulatory regimes. *The Accounting Review* 84 (3): 839–867.
- Institute of Internal Auditors (IIA). 2002. *Practice Advisory 1110-2: Chief Audit Executive (CAE) Reporting Lines*. Available at: http://som.csudh.edu/depts/adjunct/mabohebeish/ACC%20336/PA%20ADVISORY/Practice_Advisory_1110-2.pdf.
- . 2004. *Internal Auditing's Role in Sections 302 and 404 of the Sarbanes-Oxley Act*. Available at: <http://www.theiia.org/iaa/download.cfm?file=1655>.
- . 2008. *International Standards for the Professional Practice of Internal Auditing*. Available at: <http://www.theiia.org/guidance/standards-and-guidance/ippf/standards/>.
- . 2009a. *Formulating and Expressing Internal Audit Opinions*. Available at: <http://www.theiia.org/guidance/standards-and-guidance/ippf/practice-guides/>.
- . 2009b. *Practice Advisory 1110-1: Organizational Independence*. Available at: http://som.csudh.edu/depts/adjunct/mabohebeish/ACC%20336/PA%20ADVISORY/Practice_Advisory_1110-1.pdf.
- Krishnan, J. 2005. Audit committee quality and internal control: An empirical analysis. *The Accounting Review* 80 (2): 649–675.
- Krishnan, G. V., and G. Vishwanathan. 2007. Reporting internal control deficiencies in the post-Sarbanes-Oxley Era: The role of auditors and corporate governance. *International Journal of Auditing* 11: 73–90.
- Kinney, W. 2000. Research opportunities in internal control quality and quality assurance. *Auditing: A Journal of Practice & Theory* 19 (Winter): 83–90.
- Maletta, M. J. 1993. An examination of auditors' decisions to use internal auditors as assistants: The effect of inherent risk. *Contemporary Accounting Research* 9 (1): 508–525.
- McDaniel, L. S. 1990. The effects of time pressure and audit program structure on audit performance. *Journal of Accounting Research* 28 (2): 267–285.
- Messier, W. F., Jr., and A. Schneider. 1988. A hierarchical approach to the external auditors' evaluation of the internal auditing function. *Contemporary Accounting Research* 4 (2): 337–353.
- New York Stock Exchange (NYSE). 2009. *Listed Company Manual*. Available at: <http://nysemanual.nyse.com/lcm/>.
- PricewaterhouseCoopers (PwC). 2006. *2006 State of the Auditing Profession Study: Continuous auditing gains momentum*. Available at: <http://www.pwc.com/us/en/internal-audit/publications/2009-study-internal-audit-profession.jhtml>.
- Public Company Accounting Oversight Board (PCAOB). 2004. *An Audit of Internal Control Over Financial*

- Reporting Performed in Conjunction With an Audit of Financial Statements*. Auditing Standard No. 2. Washington, D.C. Available at: http://www.pcaobus.org/Rules/Rules_of_the_Board/Auditing_Standard_2.pdf.
- . 2005. *Report on the Initial Implementation of Auditing Standard No. 2, An Audit of Internal Control Over Financial Reporting Performed in Conjunction with an Audit of Financial Statements*. Release No. 2005-023. November 30. Washington, D.C. Available at: <http://pcaobus.org/Inspections/Pages/PublicReports.aspx>.
- . 2007a. *An Audit of Internal Control Over Financial Reporting That Is Integrated with An Audit of Financial Statements*. Auditing Standard No. 5. Washington, D.C. Available at: http://www.pcaobus.org/Standards/Auditing/Pages/Auditing_Standard_5.aspx.
- . 2007b. *Report on the Second-Year Implementation of Auditing Standard No. 2, An Audit of Internal Control Over Financial Reporting Performed in Conjunction with an Audit of Financial Statements*. Release No. 2007-004. April 18. Washington, D.C. Available at: <http://pcaobus.org/Inspections/Pages/PublicReports.aspx>.
- Prawitt, D. F., J. L. Smith, and D. A. Wood. 2009. Internal audit function quality and earnings management. *The Accounting Review* 84 (4): 1255–1280.
- Rogers, J. L., and P. C. Stocken. 2005. Credibility of management forecasts. *The Accounting Review* 80 (4): 1233–1260.
- Schneider, A. 1984. Modeling external auditors' evaluations of internal auditing. *Journal of Accounting Research* 22 (2): 657–678.
- . 1985. The reliance of external auditors on the internal audit function. *Journal of Accounting Research* 23 (2): 911–919.
- , and N. Wilner. 1990. A test of audit deterrent to financial reporting irregularities using the randomized response technique. *The Accounting Review* 65 (3): 668–681.
- Schwartz, S. T., and R. A. Young. 2002. A laboratory investigation of verification and reputation formation in a repeated joint investment setting. *Contemporary Accounting Research* 19 (2): 311–342.
- Securities and Exchange Commission (SEC). 2007. *Commission Guidance Regarding Management's Report on Internal Control over Financial Reporting Under Section 13(a) or 15(d) of the Securities Exchange Act of 1934*. Release Nos. 33-8810, 34-55929. Washington, D.C.: SEC.
- Shumway, T. 2001. Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business* 74 (1): 101–125.
- Stephens, N. M. 2009. Corporate governance quality and internal control reporting under SOX Section 302. Working paper, Utah State University.
- U.S. Congress. 2002. The Sarbanes-Oxley Act of 2002. Public Law 107–2004. H.R. 3763. Washington, D.C.: Government Printing Office.
- Wallace, W., and R. Kreutzfeldt. 1991. Distinctive characteristics of entities with an internal audit department and the association of the quality of such departments with errors. *Contemporary Accounting Research* 7 (2): 485–512.
- Wright, A., and R. H. Ashton. 1989. Identifying audit adjustments with attention-directing procedures. *The Accounting Review* 64 (4): 710–728.

The Causes and Consequences of Internal Control Problems in Nonprofit Organizations

Christine Petrovits

The George Washington University

Catherine Shakespeare

University of Michigan

Aimee Shih

New York University



ABSTRACT: This study examines the causes and consequences of internal control deficiencies in the nonprofit sector using a sample of 27,495 public charities from 1999 to 2007. We first document that the likelihood of reporting an internal control problem increases for nonprofit organizations that are in poor financial health, growing, more complex, and/or smaller. We then present evidence that the disclosure of weak internal controls over financial reporting is negatively associated with subsequent donor support received after controlling for the current level of donor support and other factors influencing donations. We likewise report a negative association between internal control problems and subsequent government grants. Our results suggest that donors and government agencies, important sources of capital for nonprofit organizations, react either directly or indirectly to internal control information.

Keywords: *internal control; nonprofit organizations; donors; government grants; Sarbanes-Oxley.*

Data Availability: *Data are available from sources cited in the text.*

I. INTRODUCTION

The nonprofit sector represents a sizable slice of the United States economy. Nonprofit organizations had over \$3.4 trillion in assets under their control and charitable giving to these organizations reached an estimated \$295 billion, or 2.2 percent of gross domestic product, in 2006 (Wing et al. 2008). Several recent financial scandals have highlighted the sig-

We thank Phomika Palmer and Tom Pollak at the National Center for Charitable Statistics for help with data and Guidestar for providing access to historical Form 990s. We also thank Bill Baber, Mary Billings, Teresa Gordon, Angela Gore, Erica Harris, Steve Kachelmeier (editor), Irene Kim, Sarah McVay, Dan Tinkelman, Thomas Vermeer, Michelle Yetman, workshop participants at the 2007 ARNOVA Conference, the 2008 European Accounting Association Annual Congress, the 2008 American Accounting Association Annual Meeting, the 2009 FEA Conference, University of California, Irvine, Johns Hopkins University, Georgetown University, The George Washington University, and two anonymous reviewers for helpful comments.

Editor's note: Accepted by Steven Kachelmeier

Submitted: April 2009

Accepted: July 2010

Published Online: January 2011

nificant fiduciary responsibilities of nonprofit managers as well as the relatively weak regulatory oversight of the nonprofit sector.¹ As a result, lawmakers have increased calls for nonprofit organizations to adopt more rigorous corporate governance practices, including improved internal control practices.

Internal control audits are not new to the nonprofit sector. Nonprofit organizations that receive federal funding have been subject to reviews of internal control since 1990. We make use of this unique setting to investigate the causes of internal control deficiencies and perhaps, more interestingly, the consequences of internal control reporting for these organizations. Specifically, we examine the characteristics of public charities that report internal control problems and the effect of such problems on subsequent contributions and government grants received.

Internal control is broadly defined as the process put in place by management to provide reasonable assurance regarding the achievement of effective and efficient operations, reliable financial reporting, and compliance with laws and regulations. Thus, results of internal control audits provide information on the level of risk that a nonprofit organization is not effectively carrying out its mission-related activities and fiduciary responsibilities. For this study, we define an internal control problem as the existence of a reportable condition over financial reporting or over compliance with federal program requirements.

We first model the probability of disclosing an internal control problem as a function of salient characteristics of nonprofit organizations using a sample of 27,495 public charities from 1999 to 2007. Our results generally suggest that nonprofit organizations that are more complex, financially distressed, smaller, and/or growing rapidly are more likely to disclose an internal control problem, consistent with prior research (Ge and McVay 2005; Keating et al. 2005; Doyle et al. 2007a; Ashbaugh-Skaife et al. 2007).

Next, we consider the consequences of disclosure of an internal control problem for nonprofit organizations. Previous research into the consequences of an internal control deficiency focuses predominately on for-profit firms' cost of equity capital, either directly or indirectly through the market's response to the announcement of an internal control problem. The results from this research are mixed. Ashbaugh-Skaife et al. (2009) report that the disclosure of an internal control problem is associated with a higher cost of capital. However, using a different specification, Ogneva et al. (2007) find no relation between internal control deficiencies and the cost of equity capital. Furthermore, there is mixed evidence on the market response to internal control problems, with Section 302 disclosures generating negative abnormal returns but Section 404 disclosures having no effect (Beneish et al. 2008; Hammersley et al. 2008).

Nonprofit organizations do not issue shares and their missions are not to maximize profit. While nonprofit managers are not accountable to shareholders, they are accountable to donors and grantors who provide an important source of capital. These donors and grantors do not have limitless resources and, therefore, nonprofit organizations must compete for funding. If an organization reports an internal control problem, donors could choose to support another organization where, presumably, the capital is used more efficiently. Therefore, disclosure of an internal control deficiency could result in lower subsequent contributions. Alternatively, unlike shareholders, donors do not ultimately benefit from a nonprofit organization's activities and, thus, are less likely to monitor the organization (Fama and Jensen 1983). Some donors may be unaware of the problem or may not care about the problem and, therefore, the disclosure of an internal control deficiency may be unrelated to subsequent contributions.

¹ These scandals include the conviction of the CEO of the United Way of America for fraud; the Ponzi scheme perpetrated by the Baptist Foundation of Arizona, which an audit by Arthur Anderson failed to detect, that resulted in the largest nonprofit bankruptcy ever; the embezzlement of funds from ACORN by the founder's brother; and the lavish spending of university money by the president of Oral Roberts University, to name a few.

We examine whether the disclosure of an internal control problem is associated with lower contributions received subsequently from donors using the Weisbrod and Dominguez (1986) model, which captures the responsiveness of donations to various economic factors. We use a two-stage estimation procedure to control for endogeneity between internal control problems and contributions received. Our results indicate that reportable conditions over financial reporting are negatively associated with future public support, even after controlling for the current level of public support and other drivers of contributions. Organizations that disclose internal control problems over financial reporting receive fewer contributions from individuals, corporations, and foundations in the subsequent year.

Next, we investigate the effect of disclosing internal controls deficiencies on subsequent contributions received from local, state, and federal government agencies. Because audits are mandated by the federal government for recipients of federal funding and the results of these audits are filed with the Federal Audit Clearinghouse, governmental agencies likely use the information contained in the audit reports as one factor in funding decisions. Our results are consistent with expectations. We report negative associations between reportable conditions over both financial reporting and federal program compliance and subsequent government contributions, after controlling for prior-year government contributions and political and economic determinants of governmental funding allocations.

This study informs the debate over whether public charities should adopt more rigorous corporate governance practices, particularly in relation to internal control. Recently, policymakers have focused attention on the perceived lack of accountability and transparency by charitable organizations. This increased scrutiny is not necessarily unwarranted due to the recent financial scandals and the size of the nonprofit sector. Recognizing that they must maintain the public's trust, charities are working together to convince policymakers that they can address their shortcomings without onerous regulations.² However, the nonprofit sector has not focused much attention on the particular issue of internal control. Opponents of increased regulation argue that most donors do not use detailed financial information to make giving decisions and that nonprofits do not have the funds to comply with burdensome rules (e.g., Irvin 2005; Mulligan 2007). Our evidence suggests that the internal control information currently produced by a subset of organizations in the nonprofit sector does affect, either directly or indirectly, both donors' and government agencies' funding decisions.

The results of this study should also interest nonprofit managers who make decisions about how to allocate scarce resources. During difficult economic times, when the demand for services is skyrocketing, it is essential that nonprofit organizations continue to attract donors and grantors. These organizations face tremendous pressure to focus resources on mission-related activities. However, Hager et al. (2004) argue that pressure from donors and watchdog groups to maximize mission-related spending and limit overhead costs to artificially low levels is detrimental in the long-run. Our results are consistent with the Hager et al. (2004) argument that underinvestment in administrative expenses (e.g., internal controls) can ultimately have negative consequences on mission-related activities. In particular, our evidence suggests that improving internal controls not only reduces the risk of monetary loss resulting from fraud or accounting error, but may also increase the organization's ability to deliver services by attracting additional funding.

Furthermore, this study contributes to the literature on the consequences of internal control reporting as it provides a more direct measure of the response to internal control problems. Prior

² The most prominent example of self-regulation is the National Panel on the Nonprofit Sector convened by the Independent Sector. This panel proposed extensive changes in nonprofit governance and oversight in a June 2005 report to Congress, "Strengthening Transparency, Governance, and Accountability of Charitable Organizations."

research examines the impact on the cost of equity capital, which could be considered a less direct measure of stakeholder response than donor contributions and government grants, in part because it is inferred from market models under some potentially strong assumptions. In this study, we measure stakeholder response to internal control problems by investigating the change in donor and government support.

Finally, understanding the effects of disclosure of internal control problems is important because auditors of nonprofit organizations adopted SAS No. 112, *Communicating Internal Control Matters Identified in an Audit*, in 2007 and its successor SAS No. 115 of the same title in 2009. These standards define the types of internal control deficiencies, provide detailed guidance on evaluating the severity of internal control deficiencies, and require auditors to communicate in writing to management and those charged with governance any deficiencies noted in an audit (*Professional Standards*, AICPA 2010, vol. 1, AU §352.01). As a result, these standards may influence public perception of nonprofit organizations. For example, PricewaterhouseCoopers (2006, 2) notes, "If an auditor identifies an internal control issue, it must be reported to trustees, granting agencies, and other regulators under new definitions and in a more public manner than before and, as a result, control deficiencies could be exposed to greater scrutiny by stakeholders." Thus, consequences from reporting an internal control deficiency during the sample period are likely to be amplified under today's standards.

The next section outlines current nonprofit regulatory oversight, with an emphasis on internal control reporting. Section III presents our hypotheses and related empirical models. Section IV describes the sample selection procedures and data. Section V reports our results. Section VI concludes.

II. BACKGROUND

The nonprofit sector is growing rapidly in size and complexity. Approximately 1.4 million nonprofit organizations operate in the United States today (Wing et al. 2008). These organizations vary significantly in terms of mission, size, and primary revenue source. The Internal Revenue Code defines over 25 categories of nonprofits, such as human service organizations, schools, health care providers, cultural institutions, community development corporations, affordable housing, and research laboratories. Nonprofits exist to provide a public benefit, and, therefore, receive preferential tax treatment and other regulatory privileges. Most nonprofit organizations are either public charities or private charitable foundations organized under Section 501(c)(3) of the Internal Revenue Code.³ Brown (2007) reports that, since 1997, the IRS has added to its master-file on average 39,465 exempt organizations per year, or 108 exempt organizations per day.

Regulatory oversight has not kept pace with the growth in the number of nonprofit organizations. Currently there are two regulatory mechanisms by which most nonprofit organizations are monitored: (1) the IRS via the organization's tax return (Form 990), which is required for all organizations receiving at least \$25,000 of public support, and (2) the nonprofit laws in the state of incorporation, which vary widely from state to state. These mechanisms have been criticized as insufficient to ensure that nonprofits meet their fiduciary obligations (Hansmann 1981; Atkinson 1998; Fishman 2003; Reiser 2005). In fact, the IRS acknowledges a lack of enforcement presence (Brown 2007). Other monitoring mechanisms do exist but vary by the type of organization and

³ Private foundations generally receive funding from a single source (i.e., a family or corporation), earn significant investment income, and make grants to other organizations. Public charities, as defined in Section 509(a), receive substantial support from the general public or government and actively conduct charitable operations. Private foundations are subject to various excise taxes and restrictions in order to ensure that they are using their resources for charitable purposes. Congress did not impose the same excise taxes and restrictions on public charities, presumably because donors hold public charities accountable.

type of funding sources (e.g., watchdog groups like the Better Business Bureau, program service contracts like Medicare, periodic program evaluations required by foundation and corporate donors, and state telemarketing reporting).

Congress passed the Sarbanes-Oxley Act of 2002 in an attempt to improve the accountability and oversight of public companies. Most of the provisions of the Sarbanes-Oxley Act do not apply to nonprofit organizations and no federal equivalent of the Act currently exists for nonprofits.⁴ Nevertheless, Sarbanes-Oxley influences attitudes about corporate governance in the nonprofit community (Ostrower 2007; Iyer and Watkins 2008). Policymakers at both the state and federal levels are considering various proposals aimed at enhancing nonprofit accountability (Fremont-Smith 2007). For example, Senator Charles Grassley (2006, 26), then Chairman of the Senate Finance committee, said:

Just as Congress has acted in the public interest to protect shareholders and workers from corporate mismanagement, so too must Congress demand transparency, accountability, and good governance from the nonprofit sector... Tightening rules and regulations governing the nonprofit sector will help repair the breach of trust that threatens to tarnish even the most reputable charities in America.

One of the main elements of Sarbanes-Oxley is management's responsibility for internal controls. Section 302 of the Act requires that chief executive and chief financial officers evaluate the design and effectiveness of internal controls on a quarterly basis and report an overall conclusion about the effectiveness of internal controls. Section 404 of the Act requires an annual audit of management's evaluation of internal controls and of the effectiveness of internal controls. Even though public charities are not subject to either Section 302 or Section 404 of Sarbanes-Oxley, similar requirements have been considered for the nonprofit sector. For example, the attorneys general in the states of New York and Massachusetts have proposed bills with provisions similar to the requirements in Section 302.

Some charities already are required to undergo internal control evaluations annually because they receive federal funding. Specifically, all organizations with federal expenditures greater than \$500,000 (\$300,000 for fiscal years ending before January 1, 2004) must have an audit conducted in accordance with Office of Management and Budget (OMB) Circular A-133 "Audits of States, Local Governments and Non-Profit Organizations."⁵ The results of these audits (Form SF-SAC) must be filed within nine months of the end of the fiscal year with the Federal Audit Clearinghouse and are publicly available.

The objective of an A-133 audit, also called a single audit, is to provide assurance that an organization receiving grants from the federal government is using the funds appropriately and is complying with all federal regulations (AICPA 2009). As part of an A-133 audit, auditors issue opinions on both the financial statements and on compliance with the provisions of the federal contracts or grants. In addition, auditors report on internal control over both financial reporting and federal program compliance.⁶ The reports on internal control identify whether there are any reportable conditions and, if so, whether any reportable conditions are material weaknesses.⁷

⁴ The two provisions of SOX that do explicitly apply to nonprofit organizations are whistle blower protection and document destruction policies.

⁵ OMB Circular A-133 was issued in 1990 under the name "Audits of Institutions of Higher Education and Other Non-Profit Institutions" and revised in 1996, 2003, and 2007. Our sample includes observations before and after the 2003 revision. The 2003 revision includes raising the audit threshold to \$500,000 and technical changes related to the determination of cognizant agency (Federal Register 68 (June 27, 2003): 38401-38402). We include year controls in our empirical tests to control for any differences across time.

⁶ Technically, auditors must test compliance requirements for "major programs." The determination of major programs takes into account risk, size, and oversight by a federal agency. See AICPA (2003) for more details.

⁷ Reportable conditions involve deficiencies in the design or operation of internal controls that could adversely affect the organization's financial reporting or its ability to administer its federal programs. Material weaknesses are reportable

Circular A-133 establishes certain conditions for determining whether a nonprofit organization qualifies as a low-risk auditee. This determination is based on numerous factors including prior-year audit results, third-party references, the level of oversight of the granting federal agency, and the inherent risk of the federal programs involved. To be considered low-risk, an organization must have been audited annually for the past two years and these prior audits must have resulted in clean opinions, no internal control deficiencies, and no audit findings. The risk determination affects the amount of auditing that is required to be performed under OMB A-133. For organizations not deemed low-risk, auditors are required to perform more testing and, thus, are more likely to uncover internal control problems.

The assessment of internal controls required by Circular A-133 is not identical to the assessment of internal controls required by Sarbanes-Oxley. The A-133 audit is not overseen by the PCAOB, but rather is performed in accordance with *Government Auditing Standards* issued by the GAO. As noted above, the scope of an A-133 audit goes beyond financial statements to include the requirements of federal grants or contracts. Although the scope is wider, A-133 audits are generally less stringent and less costly. Despite the differences from public company audits, A-133 audits do provide information on the level of risk that a nonprofit organization is not effectively carrying out its mission-related activities and fiduciary responsibilities.

III. HYPOTHESIS DEVELOPMENT AND EMPIRICAL MODELS

Determinants of Internal Control Deficiencies

We first examine the determinants of internal control deficiencies. A significant body of research examines the characteristics of publicly traded companies disclosing internal control problems. Ge and McVay (2005) find that firms disclosing material weaknesses are more complex, smaller, and less profitable than firms not disclosing material weaknesses. Coyle et al. (2007a) add that firms disclosing material weaknesses are younger, growing rapidly, or undergoing restructuring. Likewise, Ashbaugh-Skaife et al. (2007) find that firms reporting internal control deficiencies have more complex operations, greater exposure to accounting risk, fewer resources to invest in internal control, and a higher likelihood of using a dominant auditor.

Despite the extensive academic literature on internal controls in publicly traded companies, there is little research on internal control in the nonprofit sector. Keating et al. (2005) examine A-133 audit results from 1997 to 1999 using univariate tests. They find that smaller organizations and organizations not classified as low-risk (i.e., new grantees, organizations with prior problems) disclose more internal control problems. They also report that organizations with audits performed by national, large regional, and specialist firms report fewer internal control problems, which differs from the Ashbaugh-Skaife et al. (2007) auditor quality results for public companies. Keating et al. (2005) suggest that small nonprofit organizations, which are more likely to have internal control problems, select small audit firms.

We extend Keating et al. (2005) by examining a more comprehensive set of factors that may be associated with reporting internal control deficiencies in nonprofit organizations. Specifically, we model the likelihood of reporting internal control problems as a function of several internal control risk factors and audit detection variables. As discussed below, we expect organizations that are more complex, in poor financial health, smaller, new to federal funding, and/or growing rapidly to disclose more internal control deficiencies.

conditions in which the design or operation of internal controls does not reduce to a relatively low level the risk of material noncompliance with applicable grant requirements or with GAAP caused by error or fraud that may occur and not be detected in a timely manner (AICPA 2003, 104). OMB revised Circular A-133 in June 2007 to be consistent with SAS No. 112, which replaced the "reportable condition" concept with "significant deficiency." Because our sample period pre-dates SAS No. 112, we use the term reportable condition.

Public charities with diverse operations face challenges instituting internal controls across their various initiatives and divisions. We measure organizational complexity using the number of revenue sources (public support, government support, and/or program service revenue.) Organizations that receive funding from only one source generally engage in fewer types of charitable programs than organizations that receive funding from several sources. We predict that organizations with more sources of funding (*COMPLEXITY*) are more likely to have a variety of operations and, therefore, more likely to report an internal control problem.

Nonprofit organizations in poor financial health are less likely to have resources to invest in establishing strong internal controls. We use the existence of a going-concern paragraph in the opinion on the financial statements (*GOINGCONCERNRISK*) as a proxy for poor financial health. A going-concern paragraph indicates that the auditor has substantial doubt whether the organization can meet its obligations as they become due. We expect that organizations with a going-concern paragraph report more internal control deficiencies. Consistent with studies of public companies that use the existence of losses to measure financial health, we also include an indicator of whether the organization's revenues exceed its expenses (*SURPLUS*). We expect that charities with a surplus have fewer internal control problems.

Larger organizations (*SIZE*), as measured by total assets, have more resources and experience to draw on when implementing internal controls. For example, Greenlee et al. (2007) report that older and larger nonprofit organizations are more likely to have an internal audit function in place. Thus, we expect that smaller organizations disclose more internal control problems. Internal controls should change in response to organizational change as existing controls may be irrelevant or inefficient and new controls may be required. Rapidly growing organizations are often unable to adequately assess and update internal controls at the same pace at which organizational expansion occurs. We predict that change in size (*GROWTH*) is positively associated with the existence of internal control deficiencies. Similarly, organizations that receive federal funding for the first time (*NEWGRANTEE*) are less likely to have all of the internal controls systems in place to meet federal requirements.

We also investigate the effect of auditor type on the probability of reporting an internal control problem but do not make a prediction. On one hand, dominant audit firms (*BIG4*, *REGIONAL*, and *SPECIALIST*) have more training, experience, and exposure to litigation risk, all of which imply that these audit firms are more likely to discover internal control deficiencies. On the other hand, dominant audit firms may only contract with prestigious nonprofit organizations that are inherently less risky. This self-selection suggests that dominant audit firms are less likely to discover internal control problems at their nonprofit clients. In fact, Kitching (2009) finds evidence that donors behave as if dominant auditors are a signal of credibility.

As noted in Section II, auditors are required to determine whether a nonprofit organization qualifies as a low-risk auditee under OMB A-133. Because they are inherently less risky and because there is less testing involved, the likelihood of detecting an internal control problem is lower for low-risk auditees than it is for other auditees. Thus, we include an indicator variable if the organization is not deemed low-risk (*RISK*) as an important control in our model.

Based on the above discussion, we estimate the probability of disclosing an internal control deficiency as a function of organizational characteristics and audit detection variables as follows:

$$\begin{aligned} \text{Prob(ICD)} = & \beta_0 + \beta_1 \text{COMPLEXITY} + \beta_2 \text{GOINGCONCERNRISK} + \beta_3 \text{SURPLUS} + \beta_4 \ln \text{SIZE} \\ & + \beta_5 \text{GROWTH} + \beta_6 \text{RISK} + \beta_7 \text{NEWGRANTEE} + \beta_8 \text{BIG4} + \beta_9 \text{REGIONAL} \\ & + \beta_{10} \text{SPECIALIST} + \sum \gamma_i \text{INDUSTRY} + \sum \delta_i \text{YEAR}. \end{aligned} \quad (1)$$

Overall, we expect that the likelihood of reporting an internal control problem increases as a function of *COMPLEXITY*, *GOINGCONCERNRISK*, *GROWTH*, *RISK*, and *NEWGRANTEE* and

decreases as a function of *SURPLUS* and *SIZE*. The empirical specification also includes controls for industry and year.

Effect of Internal Control Deficiencies

We next examine the consequences of internal control deficiencies. Prior studies of public companies document that internal control problems are associated with equity market concerns. Specifically, firms reporting an internal control deficiency under Section 302 experience stock price declines, with the most negative returns for material weakness disclosures (Hammersley et al. 2008; Beneish et al. 2008). Evidence on the impact of internal control problems on the cost of capital for public companies is mixed. Ashbaugh-Skaife et al. (2009) find that internal control problems are associated with a higher cost of equity, while Ogneva et al. (2007), using a different specification, do not find an association. A limitation of these studies of public companies is that the cost of equity is a less direct measure of stakeholders' reactions, as it is inferred from a market model under certain strong assumptions.

There has been little consideration given to understanding the effect of internal control deficiencies in the nonprofit sector. Internal controls are established to provide assurance that operations are running efficiently and that financial reporting is reliable. We expect that nonprofit organizations with internal control problems have lower operating efficiency and produce lower quality financial reports, on average. Thus, internal control problems can influence directly or indirectly the amount of funds available to achieve the organization's mission. The source of funding takes many forms depending on the type of organization, including donor contributions, government grants, program service revenue, and/or debt financing. As discussed below, we examine the impact of internal control problems on donations (*PUBLIC SUPPORT*) and government grants (*GOV CONTRIBUTIONS*). Unlike the indirect cost of capital measure for public companies noted above, contributions by donors and government agencies to nonprofit organizations provide direct evidence of stakeholder reactions to internal control problems.

Public Support

Public support includes gifts received from individuals, trusts and estates, corporations, and foundations (*DIRECT SUPPORT*), as well as gifts received from federated fundraising agencies (*INDIRECT SUPPORT*), such as the United Way and the Combined Federal Campaign. Donors generally have less information about the quality of the nonprofit organization's output relative to government grantors, customers (who provide program service revenue), and creditors. Nevertheless, donors provide a substantial amount of support to the nonprofit sector. In the face of this information asymmetry, it is important to understand all factors, including the quality of internal control, that influence a donor's charitable giving decision in a competitive market for donations.

Several prior studies offer evidence that a public charity's operating efficiency is positively associated with the amount of donor support received (e.g., Weisbrod and Dominguez 1986; Posnett and Sandler 1989; Greenlee and Brown 1999; Tinkelman 2004; Tinkelman and Mankaney 2007).⁸ Further, Yetman (2008) reports that donors give less to organizations that overstate mission-related expenses and understate fundraising expenses, providing some support for the idea that donors can unravel low-quality financial statements. However, Tinkelman (1998) and Khumawala et al. (2005) both provide evidence that most donors do not unravel joint cost allocations made to strategically overstate mission-related activities. Overall, prior research suggests that, in

⁸ There are many slight variations in the definition of operating efficiency. In general, these operating efficiency measures attempt to capture how much the nonprofit organization spends on program-related activities (i.e., fulfilling its mission) relative to how much it spends on administrative and fundraising costs.

many but not all cases, donors use available information from the organization's Form 990 to distinguish higher quality nonprofit organizations from lower quality nonprofit organizations and make their giving decisions accordingly.

Because internal control deficiencies can signal a lack of effectiveness in providing charitable services and a higher probability of undetected misconduct, all else equal, we expect that nonprofit organizations with internal control deficiencies receive fewer subsequent contributions from the public than organizations with no internal control deficiencies. This hypothesis is based on the assumption that donors make giving decisions in order to assist in the provision of public goods and, thus, opt to give to organizations that can provide the public goods with minimum risks. While it is likely that some financially sophisticated donors (e.g., private foundations) actually obtain the publicly available A-133 audit report as part of the giving decision process, it is highly unlikely that all donors do. Even if donors do not directly learn that an organization has reported an internal control problem from the A-133 report, they may still indirectly receive information about that problem. For example, internal control problems can be associated with lower operating efficiency, which is observable on the more widely distributed Form 990. Alternatively, a donor may have lower quality interactions with an organization that has internal control problems (e.g., an internal control weakness causes donor acknowledgments not to be sent as required by the IRS).

There are reasons why the quality of an organization's internal controls may not affect public support. In particular, not all donors give in order to provide a public good. Some donors simply seek a warm glow (Andreoni 1990; Ribar and Wilhelm 2002) and, thus, internal control information, or any financial information for that matter, is irrelevant. Also, it may be too costly for donors to obtain and evaluate A-133 audit information.⁹ Finally, if the internal control audit results are not filed until nine months after year-end, the information may be stale. Therefore, it is an open empirical question as to whether internal control problems affect subsequent contributions.

We adapt the widely used Weisbrod and Dominguez (1986) approach to capture the responsiveness of donations to various economic factors. Weisbrod and Dominguez (1986) model public support as a function of conventional market variables, including price, fundraising expenses, and age:

$$\ln \text{PUBLIC SUPPORT}_t = \beta_0 + \beta_1 \ln \text{FUNDRAISING EXP}_{t-1} + \beta_2 \ln \text{PRICE}_{t-1} + \beta_3 \text{AGE}_t.$$

PRICE measures the cost to the donor of "purchasing" (i.e., contributing) one more dollar of the organization's charitable output. *PRICE* depends on the after-tax cost of giving, as well as the efficiency by which the organization generates output. Specifically, *PRICE* is defined as:

$$\text{PRICE} = \frac{(1 - T)}{1 - \left(\frac{\text{FUNDRAISING EXP} + \text{ADMINISTRATIVE EXP}}{\text{TOTAL EXPENSE}} \right)}.$$

Donors face the same marginal tax rate with respect to donations for all charitable organizations and, thus, we assume $T = 0$. Note that the denominator is equivalent to the program expense ratio (Program Expenses/Total Expenses) so when $T = 0$, *PRICE* equals the inverse of the program expense ratio (Total Expenses/Program Expenses).¹⁰ Theoretically, price should have a negative

⁹ This discussion suggests that any influence that internal control does have on public support is moderated by the level of sophistication of the organization's donor clientele. Unfortunately, it is impossible to test this supposition using archival data because nonprofit organizations do not disclose the identities of their donors in a consistent, systematic manner. See Baber et al. (2001) for a discussion of donor clienteles.

¹⁰ When defining *PRICE*, some studies scale fundraising and administrative expenses by public support (e.g., Weisbrod

influence on the level of giving. However, Bowman (2006) notes that, in prior empirical studies, results of tests examining the effect of price on public support are sensitive to model specification. The Weisbrod and Dominguez (1986) model also includes *FUNDRAISING EXP*, which represents the organization's effort to reduce information asymmetry, and *AGE*, which represents the organization's stock of goodwill. Both are expected to positively affect public support.

In order to test our expectation about the effect of internal control problems on donor contributions, we estimate the following equation:

$$\begin{aligned} \ln \text{PUBLIC SUPPORT}_t = & \beta_0 + \beta_1 \text{INTERNAL CONTROL DEFICIENCY}_{t-1} \\ & + \beta_2 \ln \text{FUNDRAISING EXP}_{t-1} + \beta_3 \ln \text{PRICE}_{t-1} + \beta_4 \text{AGE}_t \\ & + \beta_5 \ln \text{GOV CONTRIBUTIONS}_{t-1} + \beta_6 \ln \text{PROGRAM REVENUE}_{t-1} \\ & + \beta_7 \ln \text{PUBLIC SUPPORT}_{t-1} + \sum \gamma_i \text{INDUSTRY} + \sum \delta_i \text{YEAR}. \end{aligned} \quad (2)$$

In addition to the Weisbrod and Dominguez (1986) variables, we include government grants (*GOV CONTRIBUTIONS*) and program service revenue (*PROGRAM REVENUE*) in order to control for any crowding-out or crowding-in effects. Khanna and Sandler (2000), and Okten and Weisbrod (2000) provide evidence of a positive relation between public support and government grants and program service revenue, indicating a crowding-in effect.¹¹ Finally, we include prior-year public support to capture any other organization-specific factors, as well as industry and year controls. We are primarily interested in the coefficient on internal control deficiency, β_1 , and predict that the existence of an internal control problem is negatively associated with subsequent public support.¹²

Government Contributions

Government contributions represent gifts and grants from federal, state, and local governments.¹³ Similar to our expectation for public support, we hypothesize that nonprofit organizations with internal control deficiencies receive fewer government contributions than organizations with no internal control deficiencies. Given that the federal government mandates internal control reporting as part of the required A-133 audit, all else equal, the federal government should use this internal control information to make funding decisions.

Even though our prediction of a negative association between internal control problems and subsequent government contributions seems intuitive, there are questions about the actual deter-

and Dominguez 1986; Okten and Weisbrod 2000; Bowman 2006), while others scale by total expenses (e.g., Posnett and Sandler 1989; Tinkelman 1998; Khanna and Sandler 2000). That is, some studies measure the amount of charitable output relative to revenue received and some measure charitable output relative to total expenses. Our results are not affected by the choice of scale. We choose to present the results scaling by total expenses for a practical reason. In many instances, the sum of fundraising and administrative expenses exceeds public support. In these instances, $\log(\text{PRICE})$ is undefined. Thus, scaling by public support limits the number of usable observations.

¹¹ Crowding-out occurs when an increase in government support discourages donors from increasing their own contributions because need is already being met by government support. Crowding-in occurs when government support encourages donors to increase their own giving, often because government grants enhance the reputation of the nonprofit organization and because government grants are accompanied by more monitoring that provides additional assurance to donors that their funds are being used appropriately.

¹² We examine the association between internal control problems and public support received in the following year. To the extent that information from internal control audits is not available until nine months after the fiscal year-end and donors directly use this information, our tests are biased against finding results. Our approach is consistent with prior studies that examine the influence of program-spending ratios on subsequent years' giving using 990 data, where 990s are generally filed from 5 to 11 months after year-end.

¹³ It is important to note that government contributions on IRS Form 990 are distinct from government contracts or payments for service, which are included in program service revenue (e.g., Medicare payments received by a hospital are not classified as government contributions). Thus, it is common for organizations to receive no government contributions but still qualify for the A-133 audit because they earn revenue from federal contracts.

minants of government funding in the nonprofit sector. It is possible that political factors, and not the quality of the nonprofit organization as signaled by the internal control audit results, drive government contribution decisions. For example, the New York Attorney General recently initiated a probe into “pay-to-play” campaign donations made by nonprofit organizations to politicians in order for the nonprofit organizations to receive government grants (Dicker and Goldenberg 2009). In addition, government contributions include funds received from state and local governments, which may be less likely to use the A-133 audit report than the federal government.¹⁴ Finally, the federal government comprises a wide variety of federal agencies. There are 55 different federal agencies with oversight responsibilities in our sample, ranging from the CIA to the Peace Corps and from the National Science Foundation to the National Endowment for the Arts.¹⁵ These agencies have different missions and likely use internal control information differently. For these reasons, we empirically examine the link between internal control weaknesses and subsequent government grants.

We estimate a model of government contributions as a function of internal control problems and political and socio-economic factors, as follows:

$$\begin{aligned} \ln \text{GOV CONTRIBUTIONS}_t = & \beta_0 + \beta_1 \text{INTERNAL CONTROL DEFICIENCY}_{t-1} \\ & + \beta_2 \ln \text{FUNDRAISING EXP}_{t-1} + \beta_3 \text{LOBBYING}_t + \beta_4 \text{AGE}_t \\ & + \beta_5 \ln \text{PRICE}_{t-1} + \beta_6 \ln \text{PUBLIC SUPPORT}_{t-1} \\ & + \beta_7 \ln \text{PROGRAM REVENUE}_{t-1} \\ & + \beta_8 \ln \text{GOV CONTRIBUTIONS}_{t-1} + \beta_9 \text{GDP}_t + \sum \alpha_i \text{STATE} \\ & + \sum \gamma_i \text{INDUSTRY}. \end{aligned} \quad (3)$$

In order to address the political determinants of government funding, we include *LOBBYING*, which is an indicator variable that designates whether the organization incurred expenditures to influence legislation through communication with legislators or government officials.¹⁶ While lobbying may not, *per se*, result in more government funding, engaging in lobbying activities signifies a politically savvy nonprofit organization. Thus, we expect a positive association between lobbying and government contributions.

We include annual gross domestic product (*GDP*) to control for economic conditions.¹⁷ Governments should have a greater supply of funds available for gifts and grants when *GDP* is higher. Alternatively, the demand for government grants is higher during periods of low *GDP*. Thus, we make no predictions about the coefficient on *GDP*.

We also use state indicator variables as proxies for demand for government funding. One objective of government contributions is to redistribute revenue to geographic areas with eco-

¹⁴ Different states and municipalities have different levels of monitoring of their grant programs. It is not possible to identify the specific source(s) of government contributions from the Form 990. The largest portion of government contributions for most nonprofits comes from the federal government, although large, complex organizations receive contributions from numerous federal, state, and local agencies.

¹⁵ Nonprofit organizations expending more than \$50 million (\$25 million for fiscal years ending before January 1, 2004) in federal awards are assigned a cognizant agency. All other nonprofits are assigned an oversight agency. Generally, the cognizant or oversight agency is the federal agency that provides the predominant amount of federal funding. The purpose of the cognizant or oversight agency is to provide technical audit advice and act as a liaison between the nonprofit organization and other federal agencies with respect to audit issues.

¹⁶ Nonprofit organizations are not allowed to make political campaign donations but are allowed to engage in some lobbying activities, subject to certain limitations, without risking their tax-exempt status. See Treasury Regulations Section 56.4911 for more details.

¹⁷ Because we measure *GDP* on an annual basis, we cannot concurrently include year controls in the model.

nomic need. In some cases, the level of funding is established mathematically by the population served by the nonprofit organization (e.g., need-based formula grants). While nonprofit organizations do not operate exclusively in one geographic region, the state in which these organizations are headquartered can generally reflect characteristics of the populations that they serve (i.e., level of poverty). State indicator variables may also control for political factors (e.g., state representation on Congressional appropriation committees).

Andreoni and Payne (2003) find that nonprofit organizations reduce fundraising efforts when they receive government grants. If a nonprofit organization determines that government support will be slashed in the next period, either because the organization has fallen out of political favor or because of macroeconomic constraints, the nonprofit organization may increase fundraising. Thus, we expect a negative coefficient on *FUNDRAISING EXP*.

Generally, individual donors use *PRICE* as a measure of operating efficiency, while governments have more direct methods of monitoring an organization's efficiency (Khanna and Sandler 2000). Nevertheless, some government agencies, particularly at the state and local level, could use basic operating efficiency ratios to make decisions and/or *PRICE* could serve as a proxy for the more complex efficiency measures actually used by government agencies. Thus, we include *PRICE*, as defined in the previous section, in Equation (3). We also include *AGE* as an indicator of reputation, *PUBLIC SUPPORT* and *PROGRAM REVENUE* as controls for any crowding-in or crowding-out effects, prior-year *GOV CONTRIBUTIONS* to capture any other organization-specific factors, and industry controls. As with Equation (2), we are primarily interested in the coefficient on internal control deficiency, β_1 , and expect that the disclosure of an internal control problem is negatively associated with subsequent government contributions.

A final note—reporting a negative coefficient on β_1 in either Equation (2) or Equation (3) may indicate that disclosure of internal control deficiencies influences giving decisions. However, it is also possible that low levels of contributions result in inadequate resources necessary for a nonprofit organization to implement strong controls. To address endogeneity concerns, we implement the Heckman (1979) selection model. In the first stage, we use Equation (1) to estimate the likelihood of reporting an internal control deficiency and, using the parameters of this model, compute an inverse Mills ratio. In the second stage, we estimate Equation (2) and Equation (3) with the inverse Mills ratio as a control.

IV. SAMPLE SELECTION AND DATA

We obtain data on public charities from two sources: (1) the A-133 Single Audit database available from the Federal Audit Clearinghouse and (2) the IRS Form 990 databases available from the National Center for Charitable Statistics (NCCS). The A-133 data include general auditee information, the amount of federal awards expended, auditor name, type of audit performed, audit opinions, internal control information, and audit findings as reported on the Form SF-SAC.¹⁸ The IRS data include revenues, expenses, and balance sheet data as reported on the Form 990.¹⁹ All variables that we use from these databases are defined in Table 1.

¹⁸ Information on internal controls over major programs is not available in the electronic Single Audit database until 2001 when the federal government changed the format of Form SF-SAC. Information on internal controls over financial reporting is available for all years.

¹⁹ We use data from several different Form 990 databases. The Core Trend v2009a file provides organizational characteristics and basic financial statement data from 1998–2007. The DD Revenues and Expenses v2005 file provides a detailed breakdown of revenues and expense categories from 1998–2003. The SOI file provides a detailed breakdown of revenue and expense categories from 2004–2007 for a stratified random sample of firms selected by the IRS. The Core Supplemental v2009 file provides a detailed breakdown of revenues and expenses categories for 2004–2006 for organizations not covered in the SOI file. Note our sample does not include detailed data for the universe of organizations for 2007 but rather only for the sample in the SOI file. Our results are consistent when we exclude 2007 from our analysis.

TABLE 1
Variable Definitions

Variable	Definition
<i>RC_FS</i>	= An indicator variable that equals 1 if the A-133 audit noted reportable conditions in internal controls over financial reporting (SF-SAC Part II #3); otherwise 0.
<i>RC_GOV</i>	= An indicator variable that equals 1 if the A-133 audit noted reportable conditions in internal controls over major programs (SF-SAC Part III #5); otherwise 0.
<i>RC_ANY</i>	= An indicator variable that equals 1 if <i>RC_FS</i> equals 1 or <i>RC_GOV</i> equals 1; otherwise 0.
<i>PUBLIC SUPPORT</i>	= Total public support received for the fiscal year, defined as the sum of direct support and indirect support.
<i>DIRECT SUPPORT</i>	= Direct public support received for the fiscal year (Form 990 Line 1a).
<i>INDIRECT SUPPORT</i>	= Indirect public support received for the fiscal year (Form 990 Line 1b).
<i>PROGRAM REVENUE</i>	= Program service revenue, including government fees and contracts, received for the fiscal year (Form 990 Line 2).
<i>GOV CONTRIBUTIONS</i>	= Government contributions (grants) received for the fiscal year (Form 990 Line 1c).
<i>FUNDRAISING EXP</i>	= Fundraising expenses for the fiscal year (Form 990 Line 15).
<i>COMPLEXITY</i>	= Number of revenue sources included on Form 990 from 0–3 (<i>PUBLIC SUPPORT</i> , <i>GOV CONTRIBUTIONS</i> , and/or <i>PROGRAM REVENUE</i>)
<i>GOINGCONCERNRISK</i>	= An indicator variable that equals 1 if the A-133 audit includes a going-concern explanation (SF-SAC Part II #2); otherwise 0.
<i>SURPLUS</i>	= An indicator variable that equals 1 if total revenues (Form 990 Line 12) – total expenses (Form 990 Line 17) ≥ 0 ; otherwise 0.
<i>NEWGRANTEE</i>	= An indicator variable that equals 1 if the current year is the first year an organization expends federal contributions; otherwise 0.
<i>AGE</i>	= Number of years the organization has been tax-exempt (IRS RuleDate).
<i>SIZE</i>	= Beginning-of-year total assets (Form 990 Line 59a).
<i>GROWTH</i>	= The growth in assets, measured as the ratio of end-of-year total Assets (Form 990 Line 59b) to beginning-of-year total Assets (Form 990 Line 59a).
<i>RISK</i>	= An indicator variable that equals 1 if organization is classified as “not low risk” on the A-133 audit (SF-SAC Part III #3); otherwise 0.
<i>BIG4</i>	= An indicator variable that equals 1 if auditor (SF-SAC Part I #7) of the A-133 report is classified as one of the Big 6 auditors; otherwise 0.
<i>REGIONAL</i>	= An indicator variable that equals 1 if auditor (SF-SAC Part I #7) of A-133 report is classified as one of the Regional auditors; otherwise 0.

(continued on next page)

TABLE 1 (continued)

Variable	Definition
<i>SPECIALIST</i>	= An indicator variable that equals 1 if auditor (SF-SAC Part I #7) of the A-133 report is classified as one of the Specialist auditors; otherwise 0.
<i>PRICE</i>	= Total Expenses/Program Service Expense, where Total Expenses is the sum of fundraising expenses (Form 990 Line 15), management and general expenses (Form 990 Line 14), and program services expenses (Form 990 Line 13).
<i>LOBBYING</i>	= An indicator variable that equals 1 if organization reports lobbying expenditures to directly influence a legislative body (Form 990 Schedule A Line 37b); otherwise 0.

Source: IRS Form 990 from the National Center for Charitable Statistics and Form SF-SAC from the Federal Audit Clearinghouse.

Table 2, Panel A details the sample selection process. A merge of the A-133 data and the IRS Core Trend data on EIN and year results in a sample of 127,988 observations (27,495 public charities) from 1999 to 2007.²⁰ We use the Full Sample to shed light on the determinants of internal control weaknesses for a broad cross-section of organizations.

As discussed in the previous section, the second stage of our tests focuses on the influence of internal control problems on subsequent public support and government contributions. Some public charities receive only an immaterial amount of public support and/or government contributions (e.g., low-income housing projects). Thus, for the second stage, we further limit our sample to the subsets of organizations where public support or government contributions represent a nontrivial source of revenue. Specifically, following Tinkelman and Mankaney (2007), we eliminate observations with public support (government contributions) of less than \$100,000. This process results in a Public Support Sample of 47,318 observations (12,342 public charities) and a Government Sample of 65,415 observations (16,369 public charities). These limited samples are still significantly larger than samples in previous studies of internal control in the for-profit literature. Lack of necessary data further reduces sample size for specific tests.

In Table 2, Panel B, we classify observations into five main industries based on the National Taxonomy of Exempt Entities (NTEE) developed by the IRS. The five industries, which are the same industries used in Keating et al. (2005), include: Arts, Education, Health, Human Services, and Public Benefit. The remaining NTEE categories (i.e., Religion, International, Environment, and Unknown) are classified as "Other." Human Services organizations (e.g., Red Cross chapters, YMCAs) comprise approximately half of the sample, while Arts and Cultural organizations comprise the smallest fraction of the sample. Untabulated results indicate that, although Human Services are the most common type of nonprofit organizations receiving federal awards, these organizations are also the smallest as measured by total assets. Educational institutions, which

²⁰ Because the NCCS data contain some data errors we conduct the error-checking procedures recommended by the NCCS. We noted some organizations that had identical information in consecutive years. We could not universally determine which year contained the correct information and, thus, deleted all related years. This resulted in a loss of 9,735 observations. Also, as suggested by the 2006 *Guide to Using NCCS Data*, any suspicious observations were compared to full text versions of the Form 990 available at Guidestar (<http://www.guidestar.org>). A small number of corrections were made, primarily related to the units reported (i.e., the file listed \$5 instead of \$5 million). To our knowledge, any remaining errors create noise but do not systematically bias our tests.

TABLE 2
Sample Description

Panel A: Sample Selection

	<u>Full Sample</u>	<u>Public Support Sample</u>	<u>Government Sample</u>
Public operating charities reporting to IRS	2,261,486		
Organizations receiving A-133 audit	337,353		
Merge IRS and A-133 data	129,356		
Less audit periods other than "Annual"	(1,368)		
Full Sample	<u>127,988</u>		
Less public (government) support, $t_{-1} < \$100,000$		(75,389)	(54,384)
Less no public (government) support, data		(5,281)	(8,189)
Total Observations	<u>127,988</u>	<u>47,318</u>	<u>65,415</u>
Unique Organizations	27,495	12,342	16,369

Panel B: Observations by NTEE Classification

	<u>Full</u>	<u>Public Support</u>	<u>Government</u>
Arts	1,912	1,136	1,268
Education	15,574	9,435	9,688
Health	26,397	10,397	14,935
Human Services	69,320	20,788	30,667
Public Benefit	10,923	3,745	6,761
Other	<u>3,862</u>	<u>1,817</u>	<u>2,096</u>
Total Observations	127,988	47,318	65,415

Panel C: Observations by Auditor Type

	<u>Full</u>	<u>Public Support</u>	<u>Government</u>
Big 4	11,254	7,370	6,690
Regional	11,467	5,225	6,051
Specialist	18,768	6,042	8,493
Other	<u>86,499</u>	<u>28,681</u>	<u>44,181</u>
Total Observations	127,988	47,318	65,415

NTEE classifications: Arts (Major Group A), Education (Major Group B), Health (Major Groups E, F, G, H), Human Services (Major Group I, J, K, L, M, N, O, P), and Public Benefit (Major Groups R, S, T, U, V, W).

comprise approximately 12 percent of the Full Sample and 20 percent of the Public Support Sample, are significantly larger than other types of nonprofit organizations. In Table 2, Panel C, we classify observations by the type of auditor, which is based on audit firm size and experience in conducting A-133 audits, similar to Keating et al. (2005). The Big 4 category includes the largest public accounting firms during the sample period.²¹ The Regional category includes any of the next 25 largest accounting firms from *Accounting Today's* 2004 Top 100 Firms list (ranked by

²¹ This category includes any nonprofit organization audited by Deloitte & Touche, Ernst & Young, KPMG, PricewaterhouseCoopers (Coopers & Lybrand or Price Waterhouse), or Arthur Andersen.

revenues). The Specialist category includes accounting firms, not already classified as Big 4 or Regional, that conducted 100 or more A-133 audits during the sample period. The Other category includes all accounting firms not already classified as Big 4, Regional, or Specialist. The Specialists conducted approximately 15 percent of all A-133 audits in our sample, while the Big 4 and Regional firms each conducted 9 percent of the audits.²² Organizations receiving at least \$100,000 of public support are more likely to receive an audit by a Big 4 firm (15 percent of the Public Support Sample). In fact, untabulated results indicate that Big 4 firms audit the largest nonprofit organizations receiving federal funds. Using the Full Sample, the mean total assets of a Big 4 auditee is \$405.9 million, while the mean total assets of a Regional auditee is \$29.5 million. Regional firms, in turn, audit larger organizations than Specialists, which, in turn, audit larger organizations than the other accounting firms.

Table 3, Panel A provides descriptive statistics for the three samples. For the Full Sample, the mean (median) *SIZE* is \$45.8 million (\$2.4 million). The Public Support Sample is substantially larger, with a mean (median) *SIZE* of \$101.6 million (\$5.1 million). Across all three samples, organizations are relatively mature—the mean age for the Full Sample is 25 years. The median *INDIRECT SUPPORT* is \$0 for all samples, indicating that many organizations do not receive any indirect support from federated fundraising campaigns. Note all of the continuous variables except *AGE* and *GROWTH* are right-skewed. Thus, we use natural log transformations for these variables in our analysis.

During our sample period, the A-133 audit provided four indicators of an internal control problem: (1) if the organization discloses a reportable condition related to financial reporting (*RC_FS*); (2) if any reportable condition related to financial reporting constitutes a material weakness (*MW_FS*); (3) if the organization discloses a reportable condition related to compliance with federal program requirements (*RC_GOV*); and (4) if any reportable condition related to federal program compliance constitutes a material weakness (*MW_GOV*). Material weaknesses are a subset of reportable conditions, representing the more severe internal control problems. In addition to these four indicators, we also create a variable, *RC_ANY* (*MW_ANY*), which denotes the existence of a reportable condition (material weakness) over either financial reporting or federal program compliance.

Table 3, Panel B reports the frequency of each type of internal control problem. Given the relatively small number of observations with a material weakness, we focus primarily on reportable conditions. Overall, 14.86 percent of the sample discloses a reportable condition over financial reporting and 13.76 percent discloses a reportable condition of federal program compliance. Not surprisingly, there is significant overlap among organizations disclosing a financial statement internal control problem and organizations disclosing a federal program internal control problem. For comparison purposes, 14–15 percent of for-profit companies report a material weakness (Doyle et al. 2007a, Table 5).

V. RESULTS

Determinants of Internal Control Deficiencies

Table 4 reports simple correlations between our measures of internal control problems and organizational characteristics. As predicted, a going-concern paragraph in the audit opinion is positively associated with internal control deficiencies, while reporting a surplus is negatively associated with internal control deficiencies. Disclosure of internal control problems is positively associated with *GROWTH*, *RISK*, and *NEWGRANTEE* as expected. *SIZE* is negatively associated

²² These frequencies are quite different from the frequencies in the for-profit sector. For example, Ashbaugh-Skaife et al. (2007) report that the six dominant audit suppliers account for 84.7 percent of the audits of public companies.

TABLE 3
Descriptive Statistics

Panel A: Descriptive Statistics for Continuous Variables

	n	Mean	Q1	Median	Q3
Full Sample					
PUBLIC SUPPORT	111,939	2,559,273	0	78,972	642,824
GOV CONTRIBUTIONS	111,939	4,094,543	0	765,158	2,444,370
DIRECT SUPPORT	111,939	2,278,322	0	39,397	466,141
INDIRECT SUPPORT	111,939	280,953	0	0	12,628
SIZE (TOTAL ASSETS)	127,988	45,783,365	861,596	2,354,736	7,124,068
AGE	125,345	25.13	12.00	22.00	34.00
GROWTH	127,306	1.11	0.97	1.02	1.14
PROGRAM REVENUE	127,988	16,409,849	43,038	380,984	2,749,254
FUNDRAISING EXP	111,939	265,676	0	0	47,164
PRICE	110,836	1.45	1.07	1.14	1.23
COMPLEXITY	127,998	2.19	2.00	2.00	3.00
Public Support Sample					
PUBLIC SUPPORT	47,318	5,482,325	252,395	705,244	2,532,311
GOV CONTRIBUTIONS	47,318	6,039,142	422,783	1,289,011	3,539,211
DIRECT SUPPORT	47,318	4,896,227	153,939	509,695	2,068,485
INDIRECT SUPPORT	47,318	586,099	0	0	140,223
SIZE (TOTAL ASSETS)	47,318	101,562,496	1,696,555	5,060,441	24,603,302
AGE	46,683	32.52	18.00	29.00	47.00
GROWTH	47,283	1.12	0.98	1.05	1.16
PROGRAM REVENUE	47,318	32,651,096	86814	1,118,323	9,882,294
FUNDRAISING EXP	47,318	574,088	0	50,235	290,720
PRICE	47,163	1.40	1.10	1.16	1.24
COMPLEXITY	47,318	2.72	2.00	3.00	3.00

(continued on next page)

Panel A: Descriptive Statistics for Continuous Variables

	n	Mean	Q1	Median	Q3
Government Sample					
PUBLIC SUPPORT	65,415	3,499,496	10,519	177,774	999,367
GOV CONTRIBUTIONS	65,415	6,245,366	698,919	1,549,013	3,994,750
DIRECT SUPPORT	65,415	3,163,269	2,620	106,665	755,560
INDIRECT SUPPORT	65,415	336,228	0	0	41,745
SIZE (TOTAL ASSETS)	65,415	66,015,301	890,679	2,681,559	9,182,541
AGE	64,343	27.16	14.00	24.00	35.00
GROWTH	65,389	1.12	0.97	1.04	1.17
PROGRAM REVENUE	65,415	19,277,339	20,000	314,772	2,874,398
FUNDRAISING EXP	65,415	368,364	0	0	93,063
PRICE	65,150	1.33	1.08	1.14	1.23
COMPLEXITY	65,415	2.62	2.00	3.00	3.00

Panel B: Descriptive Statistics for Indicator Variables

	Full Sample	Public Support Sample	Government Sample
RC_FS	14.86%	13.78%	14.30%
RC_ANY	18.91%	18.01%	17.99%
RC_GOV	13.76%	13.05%	13.34%
MW_FS	5.25%	4.35%	4.94%
MW_ANY	5.97%	5.10%	5.61%
MW_GOV	3.86%	3.28%	3.74%
GOINGCONCERNRISK	1.16%	0.82%	0.99%
SURPLUS	58.00%	67.93%	64.97%
RISK	34.74%	30.68%	32.33%
NEWGRANTEE	9.37%	6.74%	5.79%
LOBBYING	2.80%	4.80%	3.69%

All variables are defined in Table 1.

TABLE 4
Pearson Correlation Matrix

	RC_GOV	COMPLEX	GOING CONCERN	SURPLUS	AGE	SIZE	GROWTH
RC_FS	0.537 (<0.0001)	-0.015 (<0.0001)	0.083 (<0.0001)	-0.008 (0.0063)	0.007 (0.0099)	-0.041 (<0.0001)	0.012 (<0.0001)
RC_ANY	0.799 (<0.0001)	-0.019 (<0.0001)	0.080 (<0.0001)	-0.012 (<0.0001)	0.024 (<0.0001)	0.001 (0.7876)	0.010 (<0.0001)
RC_GOV		-0.012 (<0.0001)	0.068 (<0.0001)	-0.018 (<0.0001)	0.001 (0.7949)	-0.002 (0.4835)	0.007 (0.0225)
COMPLEXITY			-0.015 (<0.0001)	0.147 (<0.0001)	0.195 (<0.0001)	0.202 (<0.0001)	0.011 (<0.0001)
GOING CONCERN				-0.059 (<0.0001)	-0.008 (0.0078)	-0.025 (<0.0001)	-0.0065 (0.0209)
SURPLUS					0.135 (<0.0001)	0.127 (<0.0001)	0.083 (<0.0001)
AGE						0.481 (<0.0001)	-0.022 (<0.0001)
SIZE							-0.015 (<0.0001)

	RISK	NEW GRANTEE	PRICE	GDP	LOBBY	PUBLIC SUP	GOV CONT
RC_FS	0.228 (<0.0001)	0.048 (<0.0001)	0.022 (<0.0001)	0.073 (<0.0001)	-0.015 (<0.0001)	-0.028 (<0.0001)	-0.010 (0.0079)
RC_ANY	0.248 (<0.0001)	0.042 (<0.0001)	0.023 (<0.0001)	0.102 (<0.0001)	-0.011 (<0.0001)	0.011 (0.0181)	-0.006 (0.1349)
RC_GOV	0.245 (<0.0001)	0.043 (<0.0001)	0.015 (<0.0001)	0.036 (<0.0001)	-0.011 (0.0007)	0.019 (0.0002)	-0.00002 (0.9960)
COMPLEXITY	0.003 (0.2354)	0.103 (<0.0001)	0.022 (<0.0001)	-0.232 (<0.0001)	0.059 (<0.0001)	0.182 (<0.0001)	0.176 (<0.0001)

(continued on next page)

	<i>RISK</i>	<i>NEW GRANTEE</i>	<i>PRICE</i>	<i>GDP</i>	<i>LOBBY</i>	<i>PUBLIC SUP</i>	<i>GOV CONT</i>
<i>GOINGCONCERNRISK</i>	0.056 (<0.0001)	0.003 (0.2345)	0.015 (<0.0001)	-0.001 (0.7380)	-0.015 (<0.0001)	-0.044 (<0.0001)	-0.012 (0.0015)
<i>SURPLUS</i>	-0.025 (<0.0001)	0.014 (<0.0001)	0.026 (<0.0001)	0.037 (<0.0001)	0.047 (<0.0001)	0.131 (<0.0001)	0.086 (<0.0001)
<i>AGE</i>	-0.105 (<0.0001)	-0.153 (<0.0001)	-0.007 (0.0343)	0.128 (<0.0001)	0.067 (<0.0001)	0.317 (<0.0001)	0.104 (<0.0001)
<i>SIZE</i>	-0.096 (<0.0001)	-0.143 (<0.0001)	0.033 (<0.0001)	0.108 (<0.0001)	0.131 (<0.0001)	0.548 (<0.0001)	0.197 (<0.0001)
<i>GROWTH</i>	0.021 (<0.0001)	0.0509 (<0.0001)	0.033 (<0.0001)	-0.005 (0.0925)	0.0026 (0.3529)	0.006 (0.1757)	0.0008 (0.8355)
<i>RISK</i>		0.234 (<0.0001)	0.034 (<0.0001)	-0.051 (<0.0001)	-0.034 (<0.0001)	-0.081 (<0.0001)	-0.065 (<0.0001)
<i>NEWGRANTEE</i>			0.079 (<0.0001)	-0.101 (<0.0001)	-0.026 (<0.0001)	-0.080 (<0.0001)	-0.107 (<0.0001)
<i>PRICE</i>				0.006 (0.0587)	0.009 (0.0044)	0.008 (0.0862)	-0.053 (<0.0001)
<i>GDP</i>					0.038 (<0.0001)	0.089 (<0.0001)	0.087 (<0.0001)
<i>LOBBY</i>						0.119 (<0.0001)	0.060 (<0.0001)

All variables are defined in Table 1. We use log form for all continuous variables except *AGE* and *GROWTH*. All variables except *PUBLIC SUPPORT* and *GOV CONTRIBUTIONS* use the full sample ($n = 127,988$). *PUBLIC SUPPORT* uses the public support sample ($n = 47,318$) and *GOV CONTRIBUTIONS* uses the government sample ($n = 65,415$). *RC_GOV* data is available for years beginning in 2001. p-values are in parentheses.

with *RC_FS*. The correlation between *COMPLEXITY* and the reportable condition measures is unexpectedly negative, which could be because *COMPLEXITY* is correlated with *SIZE*. For each of the empirical models estimated in the subsequent tables, we calculate variance inflation factors using ordinary least-squares and determine that multicollinearity is not a significant concern.

Table 5 presents the results from the first stage of our analysis. In Panel A, we estimate the probability of disclosing a reportable condition over financial reporting (*RC_FS*) as a function of organizational characteristics and audit detection factors, using a probit model for the three samples. The coefficients on *COMPLEXITY*, financial health (*GOINGCONCERNRISK* and *SURPLUS*), *SIZE*, *GROWTH*, and *RISK* have the predicted signs and are statistically significant. The coefficient on *NEWGRANTEE* is unexpectedly negative and significant. This negative coefficient results from the high association between *NEWGRANTEE* and *RISK*. When *RISK* is removed from the model, the coefficient on *NEWGRANTEE* is significantly positive.

In Table 5, Panel B, we estimate the probability of disclosing a reportable condition over either financial reporting or federal program compliance (*RC_ANY*). The results are similar to those in Panel A, except for the coefficients on *COMPLEXITY* in the Government Sample and *SIZE* in the Public Support and Government Samples. Overall, our evidence is consistent with the notion that less financially healthy and growing organizations disclose more internal control deficiencies. In addition, we provide some evidence that smaller and more complex organizations disclose more internal control problems.

The model in Table 5 also includes indicator variables for the type of auditor performing the A-133 audit. The coefficient on Big 4 auditors is reliably negative and significant, consistent with Keating et al. (2005). This result indicates that the probability of disclosing an internal control problem decreases if a Big 4 firm is used, and suggests that these audit firms selectively contract with certain high-quality nonprofit organizations. However, the coefficient on Regional firms is significantly positive, which indicates the likelihood of disclosing an internal control problem increases if a Regional audit firm is used.

We also estimate the probability of disclosing a reportable condition over federal program compliance (*RC_GOV*). The untabulated results are generally consistent with those in Table 5, with some minor exceptions. For the Full Sample, the coefficient on *COMPLEXITY* is not significant and for the Public Support Sample, the coefficients on *SIZE* and *GROWTH* are not significant. In addition, we estimate the probability of disclosing a material weakness (*MW_FS* and *MW_ANY*). Again, the results are generally consistent with those in Table 5. We find that all coefficients are statistically significant with the predicted sign, except for the coefficient on *GROWTH*, which is positive but not significant. Finally, we include the lagged internal control problem (either *RC_FS_{t-1}* or *RC_ANY_{t-1}*) and exclude *NEWGRANTEE* (because new grantees do not have prior-year internal control data). Not surprisingly, the coefficient on the lagged internal control problems is significantly positive. That is, if an organization reported an internal control weakness in the prior year, then the organization is more likely to report an internal control weakness in the current year.

Effect of Internal Control Deficiencies on Public Support

Table 6, Panel A reports the results from the second stage of our analysis. We estimate a regression of public support on the disclosure of reportable condition over financial reporting in the prior year and include the Inverse Mills ratio computed using the parameters from Table 5 for the Public Support Sample. For this table and all subsequent tables, we use Huber-White robust standard errors, where errors are clustered by organization.

In the first column ("Base") of Table 6, Panel A, we estimate the traditional Weisbrod and Dominguez (1986) model. Consistent with prior research, the coefficients on *FUNDRAISING EXP* and *AGE* are significantly positive, while the coefficient on *PRICE* is significantly negative. In

TABLE 5
Determinants of Internal Control Deficiencies

Panel A: Financial Statement Reportable Conditions

Variable	Pred. Sign	Full Sample	Public Support Sample	Government Sample
Intercept		-1.080*** (398.828)	-1.086*** (117.427)	-0.968*** (161.340)
COMPLEXITY	+	0.081*** (163.447)	0.045*** (7.455)	0.029** (5.992)
GOINGCONCERNRISK	+	0.737*** (455.994)	0.882*** (173.643)	0.801*** (238.627)
SURPLUS	-	-0.068*** (48.240)	-0.176*** (111.267)	-0.123*** (78.667)
SIZE	-	-0.029*** (91.230)	-0.014*** (6.670)	-0.025*** (34.136)
GROWTH	+	0.042*** (15.460)	0.052** (5.930)	0.042*** (7.260)
RISK	+	0.711*** (5751.825)	0.726*** (20960.732)	0.745*** (3218.747)
NEWGRANTEE	+	-0.032** (4.270)	-0.088*** (8.624)	-0.083*** (9.810)
BIG4	?	-0.382*** (316.585)	-0.367*** (150.812)	-0.349*** (142.823)
REGIONAL	?	0.280*** (346.814)	0.290*** (152.181)	0.243*** (132.915)
SPECIALIST	?	-0.145*** (109.964)	-0.009 (0.172)	-0.102*** (25.931)
Industry Indicators		Included	Included	Included
Year Indicators		Included	Included	Included
No. of Observations Used		127,236	47,281	65,364
No. of Observations with Internal Control Deficiencies		18,859	6,515	9,339
Likelihood Ratio		10356.535 <0.0001	3506.548 <0.0001	4857.398 <0.0001
Percent Concordant		71.90%	71.30%	70.90%

Panel B: Any Reportable Conditions

Variable	Pred. Sign	Full Sample	Public Support Sample	Government Sample
Intercept		-1.385*** (733.490)	-1.562*** (275.569)	-1.334*** (340.810)
COMPLEXITY	+	0.046*** (58.822)	0.046*** (8.833)	-0.010 (0.749)
GOINGCONCERNRISK	+	0.705*** (424.905)	0.815*** (149.436)	0.786*** (232.546)
SURPLUS	-	-0.081*** (75.011)	-0.148*** (87.255)	-0.120*** (83.362)
SIZE	-	-0.008*** (8.672)	0.010* (3.821)	-0.002 (0.338)

(continued on next page)

Panel B: Any Reportable Conditions

Variable	Pred. Sign	Full Sample	Public Support Sample	Government Sample
<i>GROWTH</i>	+	0.046 (19.673)	0.061*** (8.864)	0.048*** (10.339)
<i>RISK</i>	+	0.752*** (7243.628)	0.753*** (2531.856)	0.771*** (3820.811)
<i>NEWGRANTEE</i>	+	-0.032* (4.668)	-0.049* (2.893)	-0.080*** (9.469)
<i>BIG4</i>	?	-0.070*** (15.905)	-0.103*** (15.623)	-0.089*** (12.412)
<i>REGIONAL</i>	?	0.288*** (404.169)	0.308*** (192.368)	0.251*** (155.280)
<i>SPECIALIST</i>	?	-0.083*** (40.297)	-0.007 (0.096)	-0.049*** (6.993)
Industry Indicators		Included	Included	Included
Year Indicators		Included	Included	Included
No. of Observations Used		127,236	47,281	65,364
No. of Observations with Internal Control Deficiencies		23,996	8,516	11,749
Likelihood Ratio		11345.371 <0.0001	3954.150 <0.0001	5346.047 <0.0001
Percent Concordant		70.70%	70.30%	69.90%

*, **, *** Indicates statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.

All variables are defined in Table 1. We use log form for all continuous variables except *GROWTH*. Wald Chi-squared statistics are reported in parentheses.

model (1), we include an indicator variable for the existence of a reportable condition over financial reporting. The coefficient on RC_FS_{t-1} (-0.210) is significantly negative. In model (2), we add *GOV CONTRIBUTIONS*, *PROGRAM REVENUE*, and $PUBLIC\_SUPPORT_{t-1}$ which are all significantly associated with $PUBLIC\_SUPPORT_t$. Nevertheless, the coefficient on RC_FS_{t-1} is still significantly negative. Note that by including lagged public support, we are essentially estimating the *change* in public support associated with disclosing an internal control problem. Our evidence suggests that, all else equal, reporting internal control problems over financial reporting is associated with 3.8 percent less public support on average. In model (3) of Panel A, we estimate the influence of reporting the more severe material weakness over financial reporting on subsequent public support. Again, the coefficient is significantly negative (-0.056). As indicated in Table 6, Panel B, we find similar results when we use RC_ANY . The evidence in Table 6 is consistent with our hypothesis that reportable conditions related to the financial statements have a detrimental effect on subsequent support.

Next, we examine the components of *PUBLIC SUPPORT* in models (4) and (5). Both Panels A and B of Table 6 suggest that that *DIRECT SUPPORT* is negatively associated with the existence of an internal control weaknesses. The evidence for *INDIRECT SUPPORT* is mixed; Panel A reports a significantly negative coefficient, while Panel B reports an insignificant coefficient. Interestingly, the frequency of internal control problems is not statistically different for firms with

TABLE 6
The Effect of Internal Control Deficiencies on Public Support
Panel A: Financial Statement Internal Control Weaknesses

Variable	Base	PUBLIC SUPPORT			DIRECT	INDIRECT
		(1)	(2)	(3)	(4)	(5)
Intercept	13.079*** (132.60)	13.172*** (1212.15)	0.699*** (13.91)	0.692*** (13.80)	0.511*** (10.98)	0.048 (1.63)
RC_FS_{t-1}		-0.0210*** (-8.07)	-0.038*** (-3.09)		-0.034*** (-2.33)	-0.028** (-2.14)
MW_FS_{t-1}				-0.056*** (-2.58)		
$FUNDRAISING\ EXP_{t-1}$	0.134*** (39.10)	0.135*** (37.17)	0.022*** (22.49)	0.022*** (22.54)	0.026*** (21.66)	0.003*** (3.78)
$PRICE_{t-1}$	-0.151*** (-4.72)	-0.226*** (-4.30)	-0.058 (-1.46)	-0.059 (-1.47)	-0.031 (-0.73)	-0.016 (-1.41)
AGE	0.016*** (18.88)	0.016*** (17.42)	0.002*** (10.20)	0.002*** (10.21)	0.001*** (6.56)	0.0004** (2.00)
$GOV\ CONTRIBUTIONS_{t-1}$			0.009*** (9.74)	0.009*** (9.76)	0.009*** (8.57)	0.002** (1.95)
$PROGRAM\ REVENUE_{t-1}$			0.010*** (8.91)	0.009*** (9.76)	0.006*** (5.23)	0.001 (0.91)
$PUBLIC\ SUPPORT_{t-1}$			0.922*** (266.21)	0.922*** (266.20)		
$DIRECT\ SUPPORT_{t-1}$					0.933*** (309.96)	
$INDIRECT\ SUPPORT_{t-1}$						0.985*** (1168.79)
Inverse Mills		0.064*** (4.87)	0.006 (0.91)	0.008 (0.58)	0.010 (1.15)	-0.006 (-0.83)
Industry Indicators	Included	Included	Included	Included	Included	Included
Year Indicators	Included	Included	Included	Included	Included	Included

(continued on next page)

Panel A: Financial Statement Internal Control Weaknesses

Variable	PUBLIC SUPPORT			DIRECT		INDIRECT	
	Base	(1)	(2)	(3)	(4)	(5)	(5)
Number of Observations	44,353	40,030	37,878	37,878	37,588	36,601	36,601
R ²	35.59%	36.40%	83.20%	83.19%	90.13%	97.77%	97.77%

Panel B: Any Internal Control Weaknesses

Variable	PUBLIC SUPPORT			DIRECT		INDIRECT	
	Base	(1)	(2)	(3)	(4)	(5)	(5)
Intercept	13.079*** (132.60)	13.146*** (121.80)	0.693*** (13.83)	0.690*** (13.77)	0.509*** (10.97)	0.038 (1.31)	0.038 (1.31)
RC_ANY _{t-1}		-0.086*** (-3.77)	-0.028*** (-2.50)		-0.027*** (-2.17)	-0.018 (-1.50)	-0.018 (-1.50)
MW_ANY _{t-1}				-0.046*** (-2.29)			
FUNDRAISING EXP _{t-1}	0.134*** (39.10)	0.136*** (37.26)	0.022*** (22.51)	0.022*** (22.54)	0.026*** (21.65)	0.003*** (3.93)	0.003*** (3.93)
PRICE _{t-1}	-0.151*** (-4.72)	-0.227*** (-4.25)	-0.058 (-1.46)	-0.059 (-1.46)	-0.030 (-0.72)	-0.016 (-1.38)	-0.016 (-1.38)
AGE	0.016*** (18.88)	0.016*** (17.47)	0.002*** (10.20)	0.002*** (10.21)	0.001*** (6.59)	0.0005*** (2.10)	0.0005*** (2.10)
GOV CONTRIBUTIONS _{t-1}			0.009*** (9.69)	0.009*** (9.74)	0.009*** (8.51)	0.002*** (2.03)	0.002*** (2.03)
PROGRAM REVENUE _{t-1}			0.010*** (8.97)	0.010*** (8.94)	0.006*** (5.25)	0.001 (0.85)	0.001 (0.85)
PUBLIC SUPPORT _{t-1}			0.922*** (266.52)	0.922*** (266.27)			
DIRECT SUPPORT _{t-1}					0.934*** (310.24)		

(continued on next page)

Panel B: Any Internal Control Weaknesses

Variable	PUBLIC SUPPORT			DIRECT		INDIRECT	
	Base	(1)	(2)	(3)	(4)	(5)	
<i>INDIRECT SUPPORT_{t-1}</i>							
Inverse Mills		0.036*** (2.97)	0.008 (1.23)	0.008 (0.83)	0.009 (1.23)	0.985*** (1171.170)	
Industry Indicators	Included	Included	Included	Included	Included		
Year Indicators	Included	Included	Included	Included	Included		
Number of Observations	44,353	40,028	37,877	37,878	37,587	36,600	
R ²	35.59%	36.32%	83.21%	83.19%	90.14%	97.77%	

*, **, *** Indicates statistical significance at the 0.10, 0.05, or 0.01 level, respectively. All variables are defined in Table 1. We use log form for all continuous variables except AGE. Influential observations, identified as studentized residuals greater than 3, are removed. t-statistics are reported in parentheses. We use Huber-White robust standard errors clustered by organization.

direct support compared to firms with indirect support.²³ Those organizing and contributing to federated fund-raising campaigns do not appear to consider internal control problems as part of the giving decision in a more sophisticated manner than other donors.

We also estimate the effect of disclosing a reportable condition over federal program compliance on subsequent support (untabulated). The coefficient on *RC_GOV* is negative but not significant ($p = 0.16$). This result suggests that giving decisions made by individuals, corporations, and foundations are highly associated with internal control problems over financial reporting but not with internal control problems over federal program compliance.

We conduct a series of robustness tests on model (2) in Table 6, Panel A and Panel B. We control for macroeconomic factors by including GDP and state indicator variables. We address the possibility that low financial reserves are associated with both public support and the existence of an internal control problem by including *GOINGCONCERNRISK*. We include *SIZE*, as well as scale all dollar-denominated continuous variables by total revenue. Finally, we change the sample selection criteria to include all organizations with public support over \$1,000, instead of \$100,000. For each of these tests, we continue to find negative and significant coefficients on *RC_FS* and *RC_ANY*, suggesting donors directly or indirectly use internal control information in their giving decisions.

Finally, we investigate the association between internal control problems and subsequent public support across the six NTEE industries listed in Table 1. We find negative and significant coefficients on both *RC_FS* and *RC_ANY* for Education and Health. The coefficients on internal controls problems for the remaining industries are insignificant. According to Wing et al. (2008), excluding Religion,²⁴ Education receives the largest percentage of total public support, followed next by Health. Thus, it appears that internal control problems influence industries that receive substantial contributions from donors.

Effect of Internal Control Deficiencies on Government Contributions

Table 7 reports the results from estimating a regression of government contributions on *RC_FS*, *RC_ANY*, and *RC_GOV* for the Government Sample. We again include the inverse Mills ratio computed using the parameters from Table 5. The coefficients on *RC_FS* (-0.017) and *RC_ANY* (-0.017) are negative and significant. Unlike with public support, the coefficient on *RC_GOV* (-0.021) is also significantly negative. These results suggest that government agencies do use information regarding internal controls over federal program compliance to make funding decisions. When we use material weaknesses instead of reportable conditions, only the coefficient on *MW_GOV* is significant.

In addition, the coefficient on *FUNDRAISING_EXP* is significantly negative, consistent with Andreoni and Payne (2003). The coefficient on *LOBBYING* is positive, suggesting that organizations that engage in activities to influence legislation receive more government grants. The coefficient on GDP is negative. Overall, government contributions to nonprofit organizations appear to be a function of political and economic factors, as well as the organization's perceived ability to fulfill its mission as signaled by the results of its internal control audit.

The results in Table 7 are consistent using alternative specifications of Equation (3). In

²³ In fact, 20.7 percent of organizations that receive only indirect support disclose an internal control problem, while 19.1 percent of organizations that receive only direct support disclose a problem. Of the organizations that receive both direct and indirect support, 16.8 percent disclose an internal control problem. These percentages suggest that any differential effects of internal controls problems on direct and indirect support are not caused by superior selection methods of federated fundraising campaigns (i.e., these campaigns do not select out of giving to organizations with problems).

²⁴ Religion receives the largest amount of public support. Most religious organizations are not included in our sample because religious institutions are exempt from the Form 990 filing requirement and most do not receive federal funding.

TABLE 7
The Effect of Internal Control Deficiencies on Government Contributions

Variable	GOV CONTRIBUTIONS				
Intercept	0.561*** (10.39)	0.558*** (10.33)	0.587*** (12.45)	0.556*** (10.30)	0.588*** (10.25)
RC_FS _{t-1}	-0.017*** (-2.42)				
RC_ANY _{t-1}		-0.017*** (-2.61)			
RC_GOV _{t-1}			-0.021** (-2.52)		
MW_FS _{t-1}				-0.013 (-1.18)	
MW_ANY _{t-1}					-0.013 (-1.19)
MW_GOV _{t-1}					-0.028* (-1.86)
FUNDRAISING EXP _{t-1}	-0.001* (-1.90)	-0.001* (-1.91)	-0.0005 (-0.90)	-0.001* (-1.88)	-0.001 (-0.89)
PRICE _{t-1}	-0.026* (-1.76)	-0.025* (-1.75)	-0.024 (-1.46)	-0.026* (-1.76)	-0.024 (-1.46)
LOBBYING _t	0.018** (2.07)	0.019** (2.06)	0.020* (1.83)	0.018** (2.04)	0.019* (2.05)
AGE	0.000 (-0.16)	0.000 (-0.14)	0.000 (-0.43)	0.000 (-0.12)	0.000 (-0.43)
PUBLIC SUPPORT _{t-1}	0.005*** (8.72)	0.005*** (8.73)	0.004*** (6.29)	0.005*** (8.74)	0.004*** (6.28)
PROGRAM REVENUE _{t-1}	-0.002*** (-3.66)	-0.002*** (-3.64)	-0.002*** (-2.73)	-0.002*** (-3.63)	-0.002*** (-2.72)
GOV CONTRIBUTIONS _{t-1}	0.978*** (518.71)	0.978*** (518.82)	0.977*** (454.62)	0.978*** (518.72)	0.976*** (554.35)

(continued on next page)

TABLE 7 (continued)

Variable	GOV CONTRIBUTIONS					
	-0.014*** (-9.34)	-0.013*** (-9.19)	-0.016*** (-8.29)	-0.014*** (-9.28)	-0.014*** (-9.26)	-0.016*** (-8.47)
GDP						
Inverse Mills	0.013*** (3.03)	0.011*** (3.00)	0.012*** (2.98)	0.008 (1.50)	0.010* (1.82)	0.015*** (2.20)
State Indicators	Included	Included	Included	Included	Included	Included
Industry Indicators	Included	Included	Included	Included	Included	Included
Number of Observations	52,411	52,411	40,369	52,411	52,411	40,369
R ²	89.73%	89.73%	89.71%	89.73%	89.73%	89.71%

*, **, *** Indicates statistical significance at the 0.10, 0.05, or 0.01 level, respectively. All variables are defined in Table 1. We use log form for all continuous variables except AGE. Influential observations, identified as studentized residuals greater than three, are removed. t-statistics are reported in parentheses. We use Huber-White robust standard errors clustered by organization.

particular, we include *GOINGCONCERNRISK* and *SIZE*, scale the continuous variables by total revenue, and change the sample selection criteria to include all organizations with government contributions over \$1,000 instead of \$100,000. For each of these tests, we continue to find negative and significant coefficients on the reportable condition indicators, suggesting that government agencies use internal control information in their funding decisions.

In untabulated results, we next investigate the link between internal control problems and subsequent government contributions for the five most frequent oversight agencies in the Government Sample.²⁵ The most frequent oversight agencies (percentage of sample) are the Department of Health and Human Services (46 percent), the Department of Housing and Urban Development (16 percent), the Department of Education (14 percent), the Department of Agriculture (5 percent) and the Department of Justice (3 percent). No other federal agency represents more than 1 percent of the sample. The coefficients on all three measures of reportable conditions are significantly negative for organizations overseen by HUD. In addition, the coefficient on reportable conditions over financial reporting is significantly negative for organizations overseen by the Department of Agriculture. The coefficients on reportable conditions for organizations overseen by HHS, by far the most frequent oversight agency, are negative but not significant.²⁶ The coefficients on reportable coefficients for the remaining most frequent agencies are also insignificant.

VI. CONCLUSION

This study examines the causes and consequences of internal control weaknesses in nonprofit organizations. The nonprofit sector provides a useful setting to examine the effects of internal control disclosures because charitable giving by donors provides direct evidence of stakeholder reactions to such disclosures. We first document that the likelihood of reporting an internal control problem increases for nonprofit organizations that are complex, growing, smaller, and in poor financial health. We then present evidence that weak internal controls over financial reporting are negatively associated with subsequent public support and government contributions received after controlling for the current level of support and other factors influencing contributions. Thus, internal control information appears to affect, either directly or indirectly, the funders' giving decisions. Internal control reporting by nonprofits has been required for two decades. Our results are generalizable to the for-profit sector because they provide long-term evidence that stakeholders do indeed use internal control information to evaluate organizations.

Our specific results may interest several constituencies. First, the IRS and other regulators are reformulating laws in an attempt to increase public confidence in the integrity of exempt organizations. Second, donors want to make more informed charitable decisions. Third, watchdog groups promulgate standards to evaluate an organization's effectiveness in achieving its mission. These standards may encourage organizations to underinvest in infrastructure in the short-run. In fact, Hager et al. (2004) argue that pressure from donors and watchdog groups to maximize mission-related spending and limit overhead costs to artificially low levels is detrimental in the long-run.

²⁵ We define the oversight agency as the cognizant agency if a cognizant agency exists; otherwise it is the oversight agency designated in the A-133 report. The oversight agency is likely, but not necessarily, the predominant supplier of government contributions. To the extent that a nonprofit organization is complex and receives contributions from several federal agencies or receives most federal funding from contract revenue rather than from contributions, the oversight agency will be less meaningful for purposes of this test.

²⁶ Recall that Medicare and Medicaid are program service revenue and not part of our tests. One possible explanation for the lack of significant results for HHS is that the majority of HHS discretionary grant dollars are allocated to scientific research (<http://www.hhs.gov/grants/>). These are generally multiperiod grants. Thus, our tests may not properly cover the decision window for these grants.

Our evidence is consistent with the notion that short-term savings on administrative expenses (i.e., establishing internal controls) can ultimately have negative consequences on the organization's donor support and, thus, the organization's ability to deliver services.

Finally, nonprofit managers and board members need to understand the risks of failing to meet donors' and government agencies' expectations with regards to accountability. We estimate that, all else equal, organizations with internal control problems receive 3.8 percent less public support and 2.1 percent less government support. According to the National Council on Nonprofits (2009), the estimated cost of an audit for an organization with revenue of \$600,000 is \$12,000–\$20,000. Audit costs likely increase in the presence of internal control problems (Hogan and Wilkins 2008). Thus, in some cases, a cost-benefit analysis indicates an overall benefit from conducting periodic, thorough internal reviews of internal controls. If attestations of internal controls by external auditors are cost-prohibitive, then nonprofit organizations can seek in-kind support to help them improve their internal controls. For example, technology companies often donate technical support to nonprofit organizations. Similarly, other corporate donors with Sarbanes-Oxley experience can provide guidance on creating and maintaining adequate internal control systems.

Internal control in the nonprofit sector is a relatively unexplored area for researchers and there are many questions left to be addressed. We show that the disclosure of any internal control weakness is associated with future declines in public support, but leave open the question of which types of donors (i.e., foundations, corporations, or individuals) respond to the internal control information. Likewise, a more refined analysis would assist in determining which federal, state, and local agencies use the A-133 data. Finally, further research is needed to investigate how internal control weaknesses influence other aspects of a nonprofit organization's operations, including earnings management and executive compensation.

REFERENCES

- American Institute of Certified Public Accountants (AICPA). 2003. *Audits of States, Local Governments and Not-For-Profit Organizations Receiving Federal Awards*. New York, NY: American Institute of Certified Public Accountants.
- . 2009. *Government Auditing Standards and Circular A-133 Audits*. New York, NY: American Institute of Certified Public Accountants.
- . 2010. *AICPA Professional Standards*. New York, NY: American Institute of Certified Public Accountants.
- Andreoni, J. 1990. Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal* 100 (401): 464–477.
- , and A. Payne. 2003. Do government grants to private charities crowd out giving or fund-raising? *The American Economic Review* 93 (3): 792–812.
- Ashbaugh-Skaife, H., D. Collins, and W. Kinney. 2007. The discovery and reporting of internal control deficiencies prior to SOX-mandated audits. *Journal of Accounting and Economics* 44 (1–2): 166–192.
- , ———, ———, and R. LaFond. 2009. The effect of SOX internal control deficiencies on firm risk and cost of equity. *Journal of Accounting Research* 47 (1): 1–43.
- Atkinson, R. 1998. Unsettled standing: Who (else) should enforce the duties of charitable fiduciaries? *The Journal of Corporation Law* 23 (4): 655–700.
- Baber, W., A. Roberts, and G. Visvanathan. 2001. Charitable organizations' strategies and program-spending ratios. *Accounting Horizons* 15 (4): 329–343.
- Beneish, M., M. Billings, and L. Hodder. 2008. Internal control weaknesses and information uncertainty. *The Accounting Review* 83 (3): 665–703.
- Bowman, W. 2006. Should donors care about overhead costs? Do they care? *Nonprofit and Voluntary Sector Quarterly* 35 (2): 288–310.
- Brown, K. 2007. Letter to Senator Charles E. Grassley from Acting Commissioner of IRS. Dated June 28, 2007. Available at: <http://finance.senate.gov/press/Gpress/2007/prg072307a.pdf>.

- Dicker, F., and S. Goldenberg. 2009. Cuomo's pay-to-play rap for nonprofits. *New York Post* (October 23).
- Doyle, J., W. Ge, and S. McVay. 2007a. Determinants of weaknesses in internal control over financial reporting. *Journal of Accounting and Economics* 44 (1-2): 193-223.
- Fama, E., and M. Jensen. 1983. Separation of ownership and control. *The Journal of Law & Economics* 26 (2): 301-325.
- Fishman, J. 2003. Improving charitable accountability. *Maryland Law Review* 62: 218-287.
- Fremont-Smith, M. 2007. The search for greater accountability of nonprofit organizations: Recent legal developments and proposals for change. *Fordham Law Review* 76: 609-646.
- Ge, W., and S. McVay. 2005. The disclosure of material weaknesses in internal control after the Sarbanes-Oxley Act. *Accounting Horizons* 19 (3): 137-158.
- Grassley, C. 2006. Strengthening the nonprofit sector. *The Hill* (July 12): 26.
- Greenlee, J., and K. Brown. 1999. The impact of accounting information on contributions to charitable organizations. *Research in Accounting Regulation* 13: 111-125.
- , M. Fischer, T. Gordon, and E. Keating. 2007. An investigation of fraud in nonprofit organizations: Occurrences and deterrents. *Nonprofit and Voluntary Sector Quarterly* 36 (4): 676-694.
- Hager, M., T. Pollak, and K. Wing. 2004. Getting what we pay for: Low overhead limits nonprofit effectiveness. Nonprofit Overhead Cost Project Brief No. 3. Available at: <http://www.urban.org/publications/311044.html>.
- Hammersley, J., L. Myers, and C. Shakespeare. 2008. Market reactions to the disclosure of internal control weaknesses and to the characteristics of those weaknesses under Section 302 of the Sarbanes Oxley Act of 2002. *Review of Accounting Studies* 13 (1): 141-165.
- Hansmann, H. 1981. Reforming nonprofit corporation law. *University of Pennsylvania Law Review* 129: 497-623.
- Heckman, J. 1979. Sample selection bias as a specification error. *Econometrica* 47 (1): 153-161.
- Hogan, C., and M. Wilkins. 2008. Evidence on the audit risk model: Do auditors increase audit fees in the presence of internal control deficiencies. *Contemporary Accounting Research* 25 (1): 219-242.
- Irvin, R. 2005. State regulation of nonprofit organizations: Accountability regardless of outcome. *Nonprofit and Voluntary Sector Quarterly* 34 (2): 161-178.
- Iyer, V., and A. Watkins. 2008. Adoption of Sarbanes-Oxley measures by nonprofit organizations: An empirical study. *Accounting Horizons* 22 (3): 255-277.
- Keating, E., M. Fischer, T. Gordon, and J. Greenlee. 2005. The Single Audit Act: How compliant are nonprofit organizations? *Accounting and Financial Management* 17 (3): 285-309.
- Khanna, J., and T. Sandler. 2000. Partners in giving: The crowding-in effect of UK government grants. *European Economic Review* 44 (8): 1543-1556.
- Khumawala, S., L. Parsons, and T. Gordon. 2005. Assessing the quality of not-for-profit efficiency ratios: Do donors use joint cost allocation disclosures? *Journal of Accounting, Auditing & Finance* 20 (3): 287-309.
- Kitching, K. 2009. Audit value and charitable organizations. *Journal of Accounting and Public Policy* 28 (6): 510-524.
- Mulligan, L. 2007. What's good for the goose is not good for the gander: Sarbanes-Oxley-style nonprofit reforms. *Michigan Law Review* 105 (8): 1981-2009.
- National Council of Nonprofits. 2009. Special state focus: Audit thresholds. *Nonprofit Policy News* (April). Available at: <http://www.councilofnonprofits.org/?q=policynews/april2009#state>.
- Ogneva, M., K. Subramanyam, and K. Raghunandan. 2007. Internal control weakness and cost of equity: Evidence from SOX section 404 disclosures. *The Accounting Review* 82 (5): 1255-1297.
- Okten, C., and B. Weisbrod. 2000. Determinants of donations in private nonprofit markets. *Journal of Public Economics* 75 (2): 255-272.
- Ostrower, F. 2007. *Nonprofit Governance in the United States: Findings on Performance and Accountability from the First National Representative Study*. Washington, D.C.: The Urban Institute.
- Posnett, J., and T. Sandler. 1989. Demand for charity donations in private non-profit markets: The case of the U.K. *Journal of Public Economics* 40 (2): 187-200.
- PricewaterhouseCoopers. 2006. Internal control readiness: Why you need to act today. Available at: <http://www.pwc.com>.

- Reiser, D. 2005. Enron.org: Why Sarbanes-Oxley will not ensure comprehensive nonprofit accountability. *U.C. Davis Law Review* 38: 205–280.
- Ribar, D., and M. Wilhelm. 2002. Altruistic and joy-of-giving motivations in charitable behavior. *The Journal of Political Economy* 110 (2): 425–457.
- Tinkelman, D. 1998. Difference in sensitivity of financial statement users to joint cost allocations: The case of nonprofit organizations. *Journal of Accounting, Auditing & Finance* 13 (4): 377–394.
- . 2004. Using non-profit organization-level financial data to infer managers' fundraising strategies. *Journal of Public Economics* 88 (9–10): 2181–2192.
- , and K. Mankaney. 2007. When is administrative efficiency associated with charitable donations? *Nonprofit and Voluntary Sector Quarterly* 36 (1): 41–64.
- Weisbrod, B., and N. Dominguez. 1986. Demand for collective goods in private nonprofit markets: Can fundraising expenditures help overcome free-rider behavior? *Journal of Public Economics* 30 (1): 83–95.
- Wing, K., T. Pollak, and A. Blackwood. 2008. *The Nonprofit Almanac*. Washington, D.C.: Urban Institute Press.
- Yetman, M., 2008. Are donors misled by low quality financial reports? Working paper, University of California, Davis.

BOOK REVIEWS

Stephen A. Zeff, Editor

Editor's note: Two copies of books for review should be sent to the Book Review Editor: Stephen A. Zeff, Rice University, Jesse H. Jones Graduate School of Business, 6100 Main St., Houston, TX 77005. The policy of *The Accounting Review* is to publish only those reviews solicited by the Book Review Editor. Unsolicited reviews will not be accepted.

MATTHEW GILL, *Accountants' Truth: Knowledge and Ethics in the Financial World* (Oxford, U.K.: Oxford University Press, 2009, ISBN 978-0-19-954714-2, pp. x, 198).

More than two decades ago, I closed a book on the professions with the remark that “surely accounting is today far more socially important than medicine.” There were then few studies of the accounting profession and, above all, no serious theoretical analyses of accounting as the intellectual activity that both measured and normatively defined the economic phenomena at the center of modern life. In the years since, there have been a number of such studies and indeed a journal (*Accounting, Organizations and Society*) that has focused attention on such questions. Matthew Gill's *Accountants' Truth* falls in this new tradition. Because the book is written primarily for an audience in accounting rather than sociology, however, I am an awkward reviewer. But perhaps the accountant reader will find it interesting to know what strikes a sociologist about Gill's book.

The book describes the intellectual world of a very particular group: 20 accountants, snowball-sampled five each from the Big 4's London offices. They are all young men, nearly all ICAEW-qualified, nearly all within a few years of their original training. Since many potential interviewees declined to participate, the group is somewhat self-selected. Himself an ICAEW accountant, Gill interviewed these 20 in detail, including in his interview protocol a discussion of an accounting problem-scenario sent ahead. The book has six substantive chapters: a theoretical discussion of performance and truthfulness, an empirical analysis of reactions to the scenario, chapter-length discussions of technocracy, pragmatism, and professionalism, and a chapter on ethics.

Gill argues that accountants inevitably experience a tension between technocracy (the attempt to reduce accounting to mere rules, abstracting accounting from accountants) and pragmatism (the decision to bracket accounting's “truth” and to understand accounting practice through alternative metaphors such as strategy, sport, or family). The problem with rules is that rationalization in accounting simply makes accountants more and more alienated from the actual practice of accounting; rules make accountants automata, reducing professional commitment as well as ambiguity. As a result, “professionalism” becomes for these young men a kind of refuge—vacuous, but safe nonetheless—from the ethical problems that their alienated selves can see but lack the desire to rectify. “Ethics” becomes a matter of subtly shifting responsibility to the firm (“employee ethics”) or onto the law (“legal ethics”). Beyond these lies equivocation, yet another technique for handling the alienation inevitable in accounting work. Yet in the end Gill argues that this state of affairs is quite understandable, given the actual situations of his respondents, and that plans to shift it by the usual means of further rationalization and regulation are self-defeating. He devotes a final chapter to potential fixes for alienation, following the late Eliot Freidson—longtime dean of the sociologists of professions—in calling for a renewed commitment to a deeper and more comprehensive professionalism.

Although I found the book very interesting, I worry that what may seem interesting to me—the demonstrations of exactly how it is that accounting works as a particular discourse and how accounting “facts” are constructed—may be everyday common sense to the accounting reader. Conversely, the more general claims—for example, the claim that knowledge consists of a best stab at representing something about which there can be no final truth—are common sense in sociology, although they may be new to the accountant reader. The book doesn't

aim to further develop such general claims, but rather to apply them in the particular world of accounting (that's why it seems to me a book for accountants). And although its central idea—that increasing rationalization paradoxically increases alienation among accountants and may even facilitate equivocation—is quite interesting, that claim cannot (at least for a sociologist) be empirically established or even fully explored on the basis of only 20 interviews.

Such a reaction is inevitable in the sociologist reader. Discourse and construction are old news to me, and I have never thought that contemporary accountants merely discover facts, although I do know that a century ago accountants on-site counted physical inventory. What I wanted in the book was five times as many interviews, quota-sampled across a variety of types of accounting, and a much more detailed analysis of how different aspects of the discourse and practice of accounting go together differently (or similarly, if that is the case) for different kinds of accountants. That is, I wanted much more data and the analysis a sociologist would have written.

But that's a narrow disciplinary worry. The advantage of a very small sample is that it allows respondents to become full and complex individuals. Yet Gill's narrative does not always take full advantage of that complexity, both because he takes his respondents as a group emblematic of accounting in the abstract and because the organization of the book around substantive themes (rather than the complexities of individuals' adaptations) hides those very particular complexities from us. As a result, having chosen respondents very closely grouped in age, gender, professional seniority, and employer, Gill leaves us wondering how far these results generalize. Perhaps more disturbing, he tells us early on that many of these young men will leave for other occupations in short order; accountancy is for them a life stage, a preparation for something else. But given that, how can we take their "discourse" as that of accountancy in general? Gill claims that by looking at the core of the recently trained elite he will find the cutting edge of accountants' discourse. True enough, but one cannot help thinking that these are very young men, interviewed by a young man, and that their love of the "sport" of tax evasion, their appreciation for the clever "wheeze," and their cavalier attitude toward their society and its non-economic institutions may derive from a juvenile masculinity that will wear off as life teaches them a few nasty lessons.

The sample raises another issue. It is to be sure a sample of accountants, but it is also a sample of people who work for large organizations. The refuge in technocracy and pragmatism, the protective armor of employee ethics and avoidance of responsibility—these are the stuff of bureaucratic life. Indeed, Gill often invokes Robert Jackall's (1988) brilliant *Moral Mazes*, which concerns considerably older middle and upper managers of large organizations. And like Jackall's managers, Gill's accountants will be able to "outrun" whatever problems they create, in their case by leaving the occupation entirely rather than by new appointments elsewhere. One worries, then, that Gill's actual topic may be organization men, not accountants in particular.

My insistence on surroundings of course reflects my own work about the professions. I wanted to hear much more about accountants' relations with lawyers, tax advisors, business consultants, and other competitors in this work area. The book is very good on relations with clients, but other than a few words about the law, it doesn't tell us a lot about the other professions crowding the accountants, nor about those (e.g., business consultants) whom the accountants would themselves like to crowd out (a battle that has been going on at least since the 1920s, at least in the U.S.).

As someone whose theories are rooted in the historical evolution of professions over time, I also wanted more background on the history of accounting and on the historical moment (after neoliberalism and the so-called Big Bang in English legal affairs) in which these young men found themselves. Indeed, the book's argument that rationalization increases alienation cannot, ultimately, be established on the basis of purely contemporary data. It requires data over time.

Finally, it is of the nature of a book written for an accounting audience that it does not ask what, for a sociologist, would be the major theoretical questions. But nonetheless an accountant reader might wonder what those sociological questions are. For me, they involve the second-order clients or, to put it in John Dewey's term, the "public" of accounting. (A Marxist would call it the political economy of accounting.) Although Gill starts us with some general words about principals and agents, he moves very quickly into a data analysis that takes the larger social situation of his accountants for granted. But public accounting started not because clients sought it, but because investors demanded it. In the early years of English public accounting, when 80 percent to 90 percent of new joint stock companies failed annually and vast amounts of British capital were being invested in projects overseas, investors demanded public accounting as a means of knowing something real about the objects of their investments. Early English accounting knew that its principal "client" was the investor, not the firm audited. Today, of course, the situation is quite different. It is not clear for which principal accountants are today the agents, nor indeed which among the investors—smart money, dumb money—is the "user" envisioned by accountants when they decide to locate this or that exceptional item above or below the line. And of course the Inland Revenue has also emerged as a crucial audience for reports, as is the legal system that holds the ultimate power.

This complex arena in which the various consumers of balance sheets transact and argue with one another is never seriously analyzed, by either the respondents or by Gill himself. The general social setting of accounting, how it is paid for, who reads it, what are its distributional consequences for the society at large: none of these is treated as problematic, perhaps because they are so commonsensical to the presumed audience. But to a sociologist, these matters are the main event. The principal force driving the complexities and difficulties that Gill

portrays so well is the simple fact that accountants are providing information on the basis of which some people will try to make money at the expense of others. The audited firms that provide them with information, the short-run looter-investors, the long-run disattentive investors, the tax officials aiming to fund a government, the workers wondering about pension funding: all these people are trying quite reasonably to acquire money that others think is more properly theirs, or theirs for the taking, or whatever. These various people have widely varying levels of expertise and radically different time horizons in addition to their complexly conflicting interests in the present. Furthermore, they are all involved in a system that rewards those few who, like motorists who continue driving in a lane they know will be closed in a mile or so, free ride on the much larger group that obeys the rules to the letter.

It is this complex network of conflicting interests, antithetical publics, and high stakes that makes accounting what it is as a profession and as a part of the complex institutions of modern capitalism. Yet all of this is mostly absent from the book, or, perhaps better put, is simply taken for granted. That is, indeed, what most clearly marks the book as a book for accountants rather than sociologists.

I should close with some things I really liked. I suspect that many of them are old news to accountants, but they are what struck me as an outsider. I very much liked the insight that transparency can be a way of passing the buck. I like the Sennettian insight that accounting needs a public way to talk about ethical issues. I liked the discussion of "material importance," with the correlative concepts of user responsibility and *caveat lector*. I grew to like the respondents, whom I eventually started to recognize as old friends, perhaps as Gill intended by stretching their evidence out across the whole of his empirical analysis.

Gill is to be congratulated on a book that opens many of the important issues raised by accounting as a practice constitutive of the realities of modern society. But we are still waiting for a sociologically thorough theoretical and empirical analysis of those issues. This book is the starter or perhaps the fish course. The main course—perhaps it is Gill's next book—is yet to come.

REFERENCES

Jackall, R. 1988. *Moral Mazes: The World of Corporate Managers*. New York, NY: Oxford University Press.

ANDREW ABBOTT
Professor of Sociology
University of Chicago

EDITORIAL POLICY AND STYLE INFORMATION

EDITORIAL POLICY

According to the policies set by the Publications Committee (which were endorsed by the Executive Committee and were published in the *Accounting Education News*, June 1987), *The Accounting Review* "should be viewed as the premier journal for publishing articles reporting the results of accounting research and explaining and illustrating related research methodology. The scope of acceptable articles should embrace any research methodology and any accounting-related subject, as long as the articles meet the standards established for publication in the journal ... No special sections should be necessary. The primary, but not exclusive, audience should be—as it is now—academicians, graduate students, and others interested in accounting research."

The primary criterion for publication in *The Accounting Review* is the significance of the contribution an article makes to the literature. Topical areas of interest to the journal include accounting information systems, auditing and assurance services, financial accounting, management accounting, taxation, and all other areas of accounting, broadly defined. The journal is also open to all rigorous research methods.

The efficiency and effectiveness of the editorial review process is critically dependent upon the actions of both authors submitting papers and the reviewers. Authors accept the responsibility of preparing research papers at a level suitable for evaluation by independent reviewers. Such preparation, therefore, should include subjecting the manuscript to critique by colleagues and others and revising it accordingly prior to submission. The review process is not to be used as a means of obtaining feedback at early stages of developing the research.

Reviewers and editors are responsible for providing constructive and prompt evaluations of submitted research papers based on the significance of their contribution and on the rigor of analysis and presentation.

MANUSCRIPT PREPARATION AND STYLE

The Accounting Review's manuscript preparation guidelines follow *The Chicago Manual of Style* (15th ed.; University of Chicago Press). Another helpful guide to usage and style is *The Elements of Style*, by William Strunk, Jr., and E. B. White (Macmillan). Spelling follows *Webster's Collegiate Dictionary*.

FORMAT

1. All manuscripts should be formatted in 12-point font on 8 1/2 x 11" paper and should be double-spaced, except for indented quotations.
2. Manuscripts should be as concise as the subject and research method permit, generally not to exceed 7,000 words.
3. Margins should be at least one inch from top, bottom, and sides.
4. To promote anonymous review, authors should not identify themselves directly or indirectly in their papers or in experimental test instruments included with the submission. Single authors should not use the editorial "we".
5. A cover page should show the title of the paper, all authors' names, titles and affiliations, email addresses, and any acknowledgments.
6. The American Accounting Association encourages use of gender-neutral language in its publications.

Pagination: All pages, including tables, appendices and references, should be serially numbered. Major sections should be numbered in Roman numerals. Subsections should not be numbered.

Numbers: Spell out numbers from one to ten, except when used in tables and lists, and when used with mathematical, statistical, scientific, or technical units and quantities, such as distances, weights and measures. For example: *three days*; *3 kilometers*; *30 years*. All other numbers are expressed numerically.

Percentages and Decimal Fractions: In nontechnical copy use the word percent in the text.

Hyphens: Use a hyphen to join unit modifiers or to clarify usage. For example: *a cross-sectional equation*; *re-form*. See *Webster's* for correct usage.

Keywords: The abstract must be followed by at least three keywords to assist in indexing the paper and identifying qualified reviewers.

ABSTRACT/INTRODUCTION

An Abstract of about 100 words (150 maximum) should be presented on a separate page immediately preceding the text. The Abstract should concisely inform the reader of the manuscript's topic, its methods, and its findings. The Keywords statement should appear immediately below the Abstract. The text of the paper should start with a section labeled "I. Introduction," which provides more details about the paper's purpose, motivation, methodology, and findings. Both the Abstract and the Introduction should be relatively nontechnical, yet clear enough for an informed reader to understand the manuscript's contribution. The manuscript's title, but neither the author's name nor other identification designations, should appear on the Abstract page.

TABLES AND FIGURES

The author should note the following general requirements:

1. Each table and figure (graphic) should appear on a separate page and should be placed at the end of the text. Each should bear an Arabic number and a complete title indicating the exact contents of the table or figure. Tables and figures should define each variable. The titles and definitions should be sufficiently detailed to enable the reader to interpret the tables and figures without reference to the text.
2. A reference to each graphic should be made in the text.
3. The author should indicate where each graphic should be inserted in the text.
4. Graphics should be reasonably interpreted without reference to the text.
5. Source lines and notes should be included as necessary.

Equations: Equations should be numbered in parentheses flush with the right-hand margin.

DOCUMENTATION

Citations: Within-text citations are made using an author-year format. Cited works must correspond to the list of works listed in the "References" section. Authors should make an effort to include the relevant page numbers in the within-text citations.

1. In the text, works are cited as follows: author's last name and year, without comma, in parentheses. For example: one author, (Berry 2003); two authors, (Fehr and Schmidt 2003); three or more authors, (Dechow et al. 1998); more than one work cited, (Cole and Yakushiji 1984; Dechow et al. 1995; Levitt 1998); with two works by the same author(s), (Nelson 2003, 2005).
2. Unless confusion would result, do not use "p." or "pp." before page numbers. For example, (Dechow and Dichev 2002, 41–42).
3. When the reference list contains two or more works by the same author (as the only author or first of two or more authors) published in the same year, the suffix a, b, etc., is appended to the date in the within-text citations and in the "References" section. For example, (Johansson 2004a, 2004b, 2004c; Baiman and Rajan 2002a, 2002b; Dhaliwal et al. 2005a, 2005b).
4. When the author's name is mentioned in the text, it need not be repeated in the citation. For example: "Cohen et al. (2005) provide ..."
5. Citations to institutional works should use acronyms or short titles where practicable. For example: (NCFRR, The Treadway Commission 1987).
6. If the paper refers to statutes, legal treatises, or court cases, citations acceptable in law reviews should be used.

Reference List: Every manuscript must include a "References" section that contains only those works cited within the text. Each entry should contain all information necessary for unambiguous identification of the published work. Use the following formats (which follow *The Chicago Manual of Style*):

1. Arrange citations in *alphabetical order* according to the surname of the first author or the name of the institution or body responsible for the published work.
2. Use authors' initials instead of proper names.
3. For two or more authors, separate authors with a comma, including a comma before "and" (Dechow, P. M., R. Sloan, and A. Sweeney).
4. Date of publication follows the author's (authors') name(s).
5. Titles of journals or newspapers should not be abbreviated.

Sample entries are as follows:

- Baiman, S., and M. Rajan. 2002a. The role of information and opportunism in the choice of buyer-supplier relationships. *Journal of Accounting Research* 40 (2): 247–278.
- , and ———. 2002b. Incentive issues in inter-firm relationships. *Accounting, Organizations and Society* 27 (3): 213–238.
- Berry, R. 2003. Testimony before the Senate Committee on Homeland Security and Governmental Affairs Permanent Subcommittee on Investigations. November 18. Available at: <http://hsgac.senate.gov/files/111803berry.pdf>.
- Cohen, D., A. Dey, and T. Lys. 2005. The Sarbanes Oxley Act of 2002: Implications for compensation structure and risk-taking incentives of CEOs. Working paper, New York University, University of Chicago, and Northwestern University.
- Cole, R., and T. Yakushiji, eds. 1984. *The American and Japanese Auto Industries in Transition*. Ann Arbor, MI: University of Michigan.
- Dechow, P. M., R. Sloan, and A. Sweeney. 1995. Detecting earnings management. *The Accounting Review* 70 (2): 193–225.
- , S. P. Kothari, and R. L. Watts. 1998. The relation between earnings and cash flows. *Journal of Accounting and Economics* 25: 133–168.
- , and I. Dichev. 2002. The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review* 77 (Supplement): 35–59.
- Dhaliwal, D., M. Erickson, and O. Li. 2005a. Shareholder income taxes and the relation between earnings and returns. *Contemporary Accounting Research* 22: 587–616.
- , L. Krull, O. Li, and W. Moser. 2005b. Dividend taxes and implied cost of equity capital. *Journal of Accounting Research* 43: 675–708.
- Easton, P. 2003. Discussion of: The predictive value of expenses excluded from pro forma earnings. *Review of Accounting Studies* 8: 175–183.
- Engel, E., R. Hayes, and X. Wang. 2007. The Sarbanes-Oxley Act and firms' going-private decisions. *Journal of Accounting and Economics* (forthcoming).
- Financial Accounting Standards Board (FASB). 2006. *Accounting for Uncertainty in Income Taxes, and Interpretation of FASB Statement No. 109*. FASB Interpretation No. 48. Financial Accounting Series. Norwalk, CT: FASB.
- Fehr, E., and K. Schmidt. 2003. A theory of fairness, competition, and cooperation. In *Advances in Behavioral Economics*, edited by C. Camerer, G. Loewenstein, and M. Rabin, 271–296. New York, NY: Princeton University Press.
- Janis, I. L. 1982. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. Boston, MA: Houghton Mifflin.
- Levitt, A. 1998. The numbers game. Speech delivered at New York University, Center for Law and Business, September 28.
- Maggi, G. 1999. The value of commitment with imperfect observability and private information. *RAND Journal of Economics* (Winter) 30: 555–574.
- National Commission on Fraudulent Reporting (the Treadway Commission). 1987. *Report of the National Commission on Fraudulent Financial Reporting*. Washington, D.C.: NCFRR.
- Nelson, M. W. 2003. Behavioral evidence on the effects of principles- and rules-based standards. *Accounting Horizons* 17 (1): 91–104.
- . 2005. A review of experimental and archival conflicts-of-interest research in auditing. In *Conflicts of Interest: Challenges and Solutions in Business, Law, Medicine, and Public Policy*, edited by D. A. Moore, D. M. Cain, G. Loewenstein, and M. H. Bazerman. Cambridge, U.K.: Cambridge University Press.
- Rigdon, E. E., R. E. Schumacker, and W. Wothke. 1998. A comparative review of interaction and nonlinear modeling. In *Interaction and Nonlinear Effects in Structural Equation Modeling*, edited by R. E. Schumacker, and G. A. Marcoulides, 1–16. Mahwah, NJ: Erlbaum Associates.
- Schultz, E., and T. Francis. 2002. Companies profit on workers' deaths through "dead peasants" insurance. *Wall Street Journal* (April 19): 1.
- Securities and Exchange Commission (SEC). 2002. *Certification of Disclosure in Companies' Quarterly and Annual Reports*. Release Nos. 33-8124, 34-46427. Washington, D.C.: SEC.
- U.S. House of Representatives. 2002. The Sarbanes-Oxley Act of 2002. Public Law 107-204 [H. R. 3763]. Washington, D.C.: Government Printing Office.

Footnotes: Footnotes are not used for documentation. Textual footnotes should be used only for extensions and useful excursions of information that if included in the body of the text might disrupt its continuity. Footnotes should be inserted using the "footnote" or "endnote" feature of the word processing software which will automatically number the footnotes throughout the manuscript with superscript Arabic numerals.

SUBMISSION OF MANUSCRIPTS

Authors should note the following guidelines for submitting manuscripts:

1. Manuscripts currently under consideration by another journal or publisher should not be submitted. The author must state upon submission that the work is not submitted or published elsewhere.
2. For manuscripts reporting on field surveys or experiments: If the additional documentation (e.g., questionnaire, case, interview schedule) is sent as a separate file, then all information that might identify the author(s) must be deleted from the instrument.
3. Manuscripts should be submitted via email as Microsoft Word or PDF file to the Senior Editor, Steven Kachelmeier, at email address tar@mcombs.utexas.edu. Please submit separate files for (1) the manuscript's title page with identifying information (not forwarded to reviewers), (2) the manuscript with title page and all other identifying information removed, and (3) any necessary supplemental files, such as experimental instructions and/or response memoranda on invited revisions. The body or subject line of the email should also indicate the submission fee receipt code from AAA (see step 4 below).
4. A nonrefundable submission fee is required unless explicitly waived by the editor for invited revisions of previous submissions. The submission fee in U.S. funds is \$200.00 for members and \$400.00 for non-members of the AAA payable by credit card (VISA or MasterCard only). The payment form is available online at: <https://aaahq.org/AAAforms/journals/tarsubmit.cfm>. If you are unable to pay by credit card or have any questions, please contact the AAA Member Services Team at (941) 921-7747 or info@aaahq.org.
5. Revisions must be submitted within 12 months from the decision letter inviting a revision.

COMMENTS

The Accounting Review welcomes submissions of comments on previous TAR articles. Comments on articles previously published in *The Accounting Review* will generally be reviewed by two reviewers, usually including an author of the original article (to assist the editor in evaluating whether the submitted comment represents the prior article accurately) and an additional reviewer who is independent of the original article. If a comment is accepted for publication, the original author will generally be invited to reply. All other editorial requirements, as enumerated above, apply to proposed comments.

POLICY ON REPRODUCTION

An objective of *The Accounting Review* is to promote the wide dissemination of the results of systematic scholarly inquiries into the broad field of accounting.

Permission is hereby granted to reproduce any of the contents of the *Review* for use in courses of instruction, as long as the source and American Accounting Association copyright are indicated in any such reproductions.

Written application must be made to the American Accounting Association, 5717 Bessie Drive, Sarasota, FL 34233-2399, for permission to reproduce any of the contents of the *Review* for use in other than courses of instruction—e.g., inclusion in books of readings or in any other publications intended for general distribution. In consideration for the grant of permission by the *Review* in such instances, the applicant must notify the author(s) in writing of the intended use to be made of each reproduction.

Except where otherwise noted in articles, the copyright interest has been transferred to the American Accounting Association. Where the author(s) has (have) not transferred the copyright to the Association, applicants must seek permission to reproduce (for all purposes) directly from the author(s).

POLICY ON DATA AVAILABILITY

The AAA's Executive Committee policy (originally adopted in 1989, and amended in 2009) is that the objective of the Association-wide journals (*The Accounting Review*, *Accounting Horizons*, *Issues in Accounting Education*) is to provide the widest possible dissemination of knowledge based on systematic scholarly inquiries into accounting as a field of professional research, and educational activity. To fulfill this objective, authors are encouraged to make their data available for use by others in extending or replicating results reported in their articles.

JANUARY 2011 PLACEMENT ADS

The deadline for free position ads to be included in this section is two months prior to the date of publication. *The Accounting Review* is published in January, March, May, July, September, and November. Position ads in the "Placement Ads" are free with the purchase of a job posting in the AAA Career Center. For more information on how to purchase a job posting in the Career Center and receive your free position ad in *The Accounting Review*, go to the AAA website at <http://aaahq.org> and click on "Career Center," or call our office at 941-921-7747.

- 101 LOYOLA MARYMOUNT UNIVERSITY, Department of Accounting invites applicants for a full-time tenure-track faculty position at Assistant/Associate rank effective Fall 2010. Loyola Marymount, a comprehensive university in the mainstream of American Catholic higher education, seeks professionally outstanding applicants who value its mission and share its commitment to academic excellence, the education of the whole person, and the building of a just society. A doctorate in Accounting from an AACSB-accredited university is required. All candidates should demonstrate a strong commitment to scholarly research and excellence in teaching. Candidates with interest in Audit, Financial Accounting, and Cost areas are encouraged to apply. Salary is competitive and research support is available. To learn more about this position, contact Professor Mahmoud Nourayi, at (310) 338-5831, or mnourayi@lmu.edu. LMU is an Equal Opportunity Institution actively working to promote an intercultural learning community. Women and minorities are encouraged to apply. Visit <http://www.lmu.edu> for more information.
- 102 BOSTON COLLEGE, Carroll School of Management seeks applicants for a full-time, non-tenure-track Lecturer position in the Department of Accounting beginning Fall 2011. Applicants must have at least a master's degree or professional certification, excel in teaching at both the undergraduate and graduate levels, and be enthusiastic participants in university service activities. Ability to teach cost accounting, auditing, or accounting information systems is required. We offer competitive salaries and benefits, a collegial environment, an energetic faculty with close ties to both Boston and national financial and academic communities, and outstanding students. We seek individuals with an openness to diverse intellectual interests and an appreciation for Boston College's Jesuit and liberal arts traditions. Applicants should provide a curriculum vitae and recent teaching evaluations to: Billy Soo, Chair, Department of Accounting, Carroll School of Management, Boston College, Chestnut Hill, MA 02467; or Email: billy.soo@bc.edu. Applications received by January 15, 2011 will receive full consideration. The Carroll School of Management is an Equal Opportunity Employer.
- 103 YALE SCHOOL OF MANAGEMENT is seeking additional faculty members at the junior level in the area of accounting. Ph.D. (or final stages of dissertation) and demonstrated potential for high-quality research and teaching required; interdisciplinary orientation is preferred. Appointments will be made for the 2011–2012 academic year. To apply online, visit: <http://mba.yale.edu/faculty/openings.htm>. Please note that we are only accepting electronic applications this year. The deadline for receipt of all materials is November 30, 2010. Yale is an Equal Opportunity/Affirmative Action Employer and especially encourages applications from women and members of minority groups.

- 104 UNIVERSITY OF WISCONSIN—EAU CLAIRE, Department of Accounting and Finance seeks candidates for a probationary tenure-track faculty position at the rank of Assistant Professor of Accounting beginning August 22, 2010. Areas of teaching may include financial, managerial, or governmental accounting depending on qualifications. Teaching in the M.B.A. program is part of this position. Qualifications: Doctorate in accounting or related field must be received no later than May 31, 2012. Professional certification (CPA, CMA), evidence of scholarly activity and relevant work experience are all considered assets. Applicants should demonstrate capacity for high-quality teaching. Application procedure: Only electronic applications are accepted. Send letter of application, graduate school transcript, resume, summary of student evaluations (if available), and contact information for three references to: dbecker@uwec.edu; Dr. D'Arcy Becker, Chairperson, Department of Accounting and Finance, University of Wisconsin—Eau Claire, Wisconsin 54702-4004. For priority consideration, all application materials must be received by October 15, 2010; screening will continue until the position is filled. The University reserves the right to contact additional references with notice given to the candidates at an appropriate time in the process. Applicants' names are subject to public release unless confidentiality has been requested in writing. Names of all finalists will be released upon request. A criminal background check is required prior to employment. Doctorate in accounting or related field must be received no later than May 31, 2012. Professional certification (CPA, CMA), evidence of scholarly activity, and relevant work experience are all considered assets. Applicants should demonstrate capacity for high-quality teaching (as evidenced by teaching evaluations if available). UW—Eau Claire is an AA/EOE, dedicated to enhancing diversity. To learn more, visit our website: <http://www.uwec.edu/acadaff/jobs/>.
- 105 UNIVERSITY OF SOUTH CAROLINA, School of Accounting is accepting applications for a tenure-track position beginning fall 2011. Candidates should have primary teaching and research interests in international accounting, and must possess a doctorate or expect to complete it by August 2011. In addition to being committed to excellent teaching, candidates must be ambitious and motivated researchers who are committed to publishing in top tier accounting journals. Salaries are competitive and commensurate with experience and achievements. All faculty searches are subject to the availability of funding. Submit curriculum vitae, working papers, and defended dissertation proposal with data to: Professor Al Leitch, Darla Moore School of Business, University of South Carolina, Columbia, SC 29208; Email: leitch@moore.sc.edu. The University of South Carolina is an AA/EOE. Minorities and women are strongly encouraged to apply.
- 106 SAINT LOUIS UNIVERSITY, a Catholic, Jesuit institution dedicated to student learning, research, health care, and service, is seeking applicants for a tenure-track position in Accounting at the Senior Assistant/Associate/Full Professor level, beginning Fall 2011. Candidate must demonstrate a commitment to teaching excellence and quality research and publication. Service to the department and university are also required in addition to teaching and scholarship. Professional certification is desirable. Teaching interest is open and all areas will be considered except financial accounting. A Ph.D. in Accounting is required. Dissertation-stage candidates could be considered under exceptional circumstances. Compensation package is competitive. Interested candidates must apply online at: <http://jobs.slu.edu> (Req. 20100775); submit a cover letter and current curriculum vitae. All other correspondence regarding this position can be sent to: Dr. Ananth Seetharaman, Chair of Accounting, John Cook School of Business, Saint Louis University, 3674 Lindell Boulevard, St. Louis, MO 63108. Saint Louis University is an Affirmative Action/Equal Opportunity Employer, and encourages nominations of and application from women and minorities.

- 107 UNIVERSITY OF MARYLAND, COLLEGE PARK, Robert H. Smith School of Business, Department of Accounting and Information Assurance is seeking two (tenure-track) faculty positions at the Assistant Professor level to begin August 2011. For full consideration of all qualified individuals interested in the above noted position, applicants must apply online at: <http://jobs.umd.edu>; and submit a letter expressing their interest, along with a curriculum vitae, publications or job paper, and evidence of teaching ability by December 31, 2010. Applicants are expected to have a Ph.D., be able to generate research publications in top academic journals, and have demonstrated effectiveness in the classroom. While all areas of specialization will be considered, managerial accounting or auditing will be viewed positively. The University of Maryland, College Park actively subscribes to a policy of Equal Employment Opportunity, and will not discriminate against any employee or applicant because of race, age, sex, color, physical or mental handicap, national origin, or political affiliation.
- 108 SAN FRANCISCO STATE UNIVERSITY has one tenure-track faculty opening in taxation. Candidates must have a terminal degree (or a candidate may be in the final stages of dissertation completion). A Ph.D. or D.B.A. is preferred but candidates with a J.D. and an L.L.M. in taxation will also be considered. Candidates should have a strong commitment to both research and teaching. Professional experience or certification is desirable. The position will be filled at a rank commensurate with experience. The Department of Accounting offers concentrations in Accounting for B.S., M.B.A., and M.S.B.A. degrees. The College of Business is AACSB-accredited at both the undergraduate and graduate levels. Accounting is one of eight departments in the College of Business and has approximately 17 faculty and 1,200 undergraduate accounting majors. Send a letter of application, curriculum vita, teaching evaluations, and the names and telephone numbers of three references to: Dr. Jiunn Huang, Chair, Department of Accounting College of Business, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA 94132. SFSU is an Affirmative Action/Equal Opportunity employer.
- 109 CLARKSON UNIVERSITY seeks an energetic scholar for a tenure-track position at the Assistant or Associate level in Accounting, balanced between teaching and research beginning in August 2011. All areas of Accounting will be considered. The position requires a Ph.D. in Accounting, evidence of teaching effectiveness and significant research potential. The School of Business has identified Supply Chain Management, Entrepreneurship, and Environmental Studies as key strategic directions and would particularly welcome applicants with an interest or previous work in any of these areas. Clarkson provides a quality environment for research including a two-course per semester teaching load for tenure-track faculty, a collegial environment, research, and other financial and nonfinancial support. Clarkson University values a culturally diverse faculty and strongly encourages applications from women and minority candidates. The University is an Equal Opportunity/Affirmative Action Employer. For more information and to apply online, visit the University's website at: <http://www.clarkson.edu/hr>.
- 110 AUBURN UNIVERSITY MONTGOMERY invites applicants for a nine-month tenure-track position at the Assistant or Associate level in Accountancy. Position begins in Fall 2011 semester. Candidates should possess expertise and teaching experience in Tax Accounting. The successful candidate will hold a Ph.D. in accounting and experienced applicants should have a well-established record of publications and effective teaching. Expectations include commitment to excellence in teaching, research, and service. Evidence of strong instructional effectiveness and scholarly productivity is essential, as are strong communication and interpersonal skills. Business experience and professional certification are highly desirable. Ph.D. in Accounting; Experience in teaching tax; CPA certification highly desirable. Please contact Dr. Keren Deal at (334) 244-3227 or kdeal@aum.edu for more information.

- 111 CENTRAL COLLEGE seeks applications for a tenure-line position as an Assistant Professor of Accounting. Minimum qualifications include M.B.A./M.A. in accounting with CPA/CMA designation. Candidates with college teaching experience and/or holding a Ph.D. in accounting will be given preference. Please contact Marilyn Vrban at (641) 628-5175 or vrbanm@central.edu if you have any questions regarding the open position. Visit <https://www.central.edu/jobseekers> for full position description and application process. AA/EOE.
- 112 COLLEGE OF THE HOLY CROSS, Economics/Accounting Department invites applications for a tenure-track position beginning August 2011. The Department seeks to hire an accountant whose teaching and research interests might include managerial, auditing, international, or professional ethics. Ph.D. or proximate completion required. Strong candidates will demonstrate a dedication to excellent teaching within a liberal arts environment and a commitment to research, publication, and service. This position carries a 3-2 teaching load with a full-salary, one-semester research leave prior to tenure review and generous sabbatical and fellowship leaves for senior faculty. Submit cover letter, curriculum vita, three recommendation letters, research paper, statement of teaching philosophy, and undergraduate and graduate transcripts by December 15, 2010. The College of the Holy Cross is a highly selective Catholic liberal arts college in the Jesuit tradition. It enrolls about 2,900 students and is located in a medium-sized city 45 miles west of Boston. Holy Cross belongs to the Colleges of Worcester Consortium (<http://www.cowc.org>) and the New England Higher Education Recruitment Consortium (<http://www.faculty.harvard.edu/01/013.html>). Contact: Kolleen Rask, Economics Chair, Campus Box 45A, College of the Holy Cross, Worcester, MA 01610-2395. Website: <http://www.holycross.edu/departments/economics/website/>. The College complies with all Federal and Massachusetts laws concerning Equal Opportunity and Affirmative Action in the workplace.
- 113 MARIST COLLEGE, School of Management at seeks an Assistant/Associate Professor of Accounting beginning Fall 2011. This tenure-track position will involve teaching both undergraduate and graduate courses (including online courses). The business and accounting programs are fully accredited by AACSB International. Candidates should have an earned doctorate in accounting or a related field. Professional experience and certification (CPA, CMA, CIA, CISA, etc.) are also desirable. Marist College is a highly selective, independent, liberal arts institution located in the historic Hudson River Valley of New York. Marist has been recognized for excellence by U.S. News & World Report, The Princeton Review, Kiplinger's Personal Finance, Entrepreneur Magazine, and Barron's Best Buys in College Education, and is noted for its leadership in the use of technology to enhance the teaching and learning process. To learn more or to apply, please visit: <https://jobs.marist.edu>. Only online applications are accepted. EOE/AA.
- 114 THE UNIVERSITY OF NORTH CAROLINA AT GREENSBORO, Department of Accounting and Finance invites applicants for a tenured or tenure-track position starting fall 2011. The academic rank is open and the salary competitive. Greensboro is located in the Piedmont Triad region near the Research Triangle and other educational institutions. The Department consists of 13 full-time faculty and offers B.S. degrees and a M.S. degree in Accounting. The Bryan School is accredited by AACSB International in both business and accounting. The Department has subscriptions to CRSP, Compustat, and WRDS. A doctorate (or short-term ABD status) in Accounting from an accredited university is required. We will consider all areas but preference will be given to individuals with research and teaching interests in financial accounting and/or tax. For full consideration, submit a letter of interest, curriculum vita, and references to Accounting Search Committee at: ACCFIN@uncg.edu; or by mail to: UNCG, Bryan School of Business and Economics, PO Box 26165, Greensboro, NC 27402-6165. Applications will be accepted until the position is filled. AA/EOE. Moreover, UNCG is committed to recruitment and advancing women and minorities at all faculty/staff levels.

- 115 THE UNIVERSITY OF NORTH CAROLINA AT CHARLOTTE, Belk College of Business invites applications for full-time, tenure-track position(s) in Accounting. All ranks will be considered. A Ph.D. degree from an accredited school in the field of accounting or closely related discipline is required. A strong commitment to excellence in high-quality, innovative research and teaching at the undergraduate and graduate levels is expected. Candidates at the Associate or Professor level must have an outstanding research record commensurate with the level of appointment. Specializations in all accounting subfields are encouraged to apply. Teaching load and salary are competitive with research universities. The starting date is August 2011. All applications will be considered strictly confidential. The review of applications will begin immediately and applications will be accepted until the position is filled. Apply electronically at: <https://jobs.uncc.edu>. If you have any questions concerning this position(s), please contact: Professor Jack Cathey, Chair of the Search Committee, at jmcathey@uncc.edu or (704) 687-7690. Only electronic submissions will be accepted. Please attach the following documents with your electronic submission: application letter and vita. Finalists will be asked to forward official transcripts, letters of reference, and other supportive materials as requested by the search committee. The University of North Carolina at Charlotte is an Affirmative Action/Equal Opportunity Employer. Women, members of minority groups, and persons with disabilities are encouraged to apply.
- 116 LAMAR UNIVERSITY, Department of Accounting and Business Law seeks a tenure-track Assistant or Associate Professor of Accounting. Accounting information systems or tax is preferred, but not required; however, must be able to teach in several content areas at both graduate and undergraduate level. Should demonstrate a record of excellence in (or potential for) innovative teaching, quality published/publishable research, service to the university/community, as well as a desire to build a quality accounting program. Official transcripts are required at the time of employment. Positions will be filled pending available funding. For a detailed job description and application process visit: <https://jobs.lamar.edu/postings/134>. This position is security-sensitive and thereby subject to the provisions of the Texas Education Code §51.215, which authorizes the employer to obtain criminal history record information. The successful candidate must have a Ph.D. (or ABD near completion) from an AACSB-accredited program or from an institution with an equivalent reputation. Individuals with an L.L.M. in Taxation and a commitment to research will also be considered. AA/EOE/ADA. Lamar is committed to excellence through diversity.
- 117 UNIVERSITY OF MELBOURNE, Melbourne Business School (MBS), which provides the M.B.A. and E.M.B.A. programs for the University, invites applications for a faculty position in financial accounting at any rank to begin September 2011 or January 2012. MBS offers a multi-disciplinary environment, as well as interaction with colleagues at the Faculty of Business and Economics within the wider University of Melbourne community. MBS faculty regularly publish in top-tier international journals. Salary, benefits, and teaching loads are competitive with U.S. and other leading international business schools. Research support is excellent. Candidates should have or be close to completing a Ph.D. in Accounting. Consistent with rank, applicants will be evaluated on their potential for and/or record of excellence in research and teaching. Graduate teaching experience is desirable, but not required. Applicants should submit a cv, working paper, teaching evaluations (where applicable), and three reference letters (or names of three references) to: j.frederickson@mbs.edu.

- 118 BALL STATE UNIVERSITY seeks candidates for a tenure-track position as Assistant or Associate Professor of Accounting, starting August 22, 2011. Responsibilities include: teaching courses in financial, managerial, governmental and not-for-profit, international, and/or accounting information systems at the undergraduate and/or graduate levels and appropriate scholarship and service. Minimum qualifications: ABD toward doctorate in accounting or other business related field, completion of degree by November 1, 2012; evidence of scholarly research; teaching experience in accounting or other business related field at the university level. Preferred qualifications: Ph.D. or D.B.A. in accounting; full-time teaching experience at the university level; significant record of publications in refereed journals in the accounting area and/or evidence of other scholarly contributions; and CPA or CMA certificate. Competitive salary and benefits package. Send letter of application, curriculum vitae, official transcripts, and the names of three references to: Dr. Lucinda L. Van Alst, Chair, Department of Accounting, Ball State University, WB 303, Muncie, IN 47306. Review of applications will begin immediately and will continue until the position is filled. Visit <http://www.bsu.edu>. Ball State University is an Equal Opportunity/Affirmative Action Employer and is strongly and actively committed to diversity within its community.
- 119 THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL, Kenan-Flagler Business School is seeking to fill one or more tenure-track, tenured, or fixed-term faculty positions in the Accounting area starting July 1, 2011. Successful candidates will have strong research skills and be expected to publish in top-tier academic journals. Positions require a Ph.D. granted, or nearly completed, in the academic field identified, or in a related field. Hired candidates with a Ph.D. can anticipate an initial appointment of Assistant Professor, Associate Professor, or Professor. Hired candidates with a Ph.D. nearly completed can anticipate an initial appointment of Instructor. The successful candidate will be both a productive researcher, and a creative and effective teacher capable of contributing to the School's Undergraduate, M.B.A., Master's of Accounting, Ph.D., and Executive Development Programs. Qualified applicants are to complete the application process online at <http://jobs.unc.edu/2500448>, in order to be considered. Full submissions should include the following: curriculum vitae, sample research papers, example of teaching effectiveness (e.g., teaching evaluations, comments from students), and four original letters of recommendation. Please also upload letters of recommendation in the web system when applying on this recruitment. For additional information, send an email to: AccountingRecruiting@unc.edu. Original materials should be mailed to: Kenan-Flagler Business School, University of North Carolina at Chapel Hill, CB 3490, McColl Building, Chapel Hill, NC 27599-3490, Attn: Accounting. The University of North Carolina at Chapel Hill is an Equal Opportunity Employer.
- 120 FINANCIAL/ACCOUNTING/INSURANCE/TAX TRANSLATOR with potential earnings in excess of \$125,000 (€90,000) per year. An international translation agency headquartered in Paris, France, and specializing in financial and legal translations for blue-chip clients is constantly seeking to hire home-based translators and editors. Candidate should be: experienced and/or qualified accounting, finance, insurance, or tax professional; able to write clearly and concisely in your native language; capable of meeting demanding deadlines; fluent in/with good working knowledge of French (and/or one other major European language) in addition to your native language, with good comprehension of the foreign-language terminology in your subject area. Looking for a career change and the opportunity to work from home at hours that suit you? Send your resume and cover letter, or any inquiries, to: recrutement@tectrad.com.

- 121 UNIVERSITY OF NOTRE DAME, Department of Accountancy invites applications for position openings at all levels: Assistant, Associate, or Full Professor. All research areas will be considered. The appointment is expected to begin July 1, 2011. Applicants for Assistant Professor positions should expect to complete doctoral degree requirements before the start of the contract period. All candidates should have a strong commitment to scholarly research and excellence in teaching. Salary is very competitive and research support is excellent. Please submit curriculum vita, three letters of recommendation, teaching evaluations, and an example of scholarly work to: acctdept@nd.edu. Also, address inquiries to: acctdept@nd.edu. The deadline for applications is January 15, 2011. The University of Notre Dame is an Affirmative Action Employer with a strong commitment to fostering a culturally diverse atmosphere for faculty, staff, and students. Women, minorities, and those attracted to a university with a Catholic identity are especially encouraged to apply. For more information about the University of Notre Dame, the Mendoza College of Business, and the Department of Accountancy please visit the department's website at: <http://business.nd.edu/accountancy/>. University of Houston—Downtown, FACIS Department, College of Business is seeking applicants for one tenure-track position at the rank of Assistant Professor and another at the rank of Associate Professor, which, depending on credentials, can include a named professorship. While the Search Committee will accept applications until the positions are filled, interested individuals are encouraged to apply by January 1, 2011 to assure full consideration. Online applications are required at: <https://jobs.uhd.edu/applicants>. UHD enrolls over 12,000 undergraduate and graduate students and offers daytime, evening, and weekend courses through both classroom and various technology-enabled mediums. UHD has received top rankings by U.S. News & World Report for wireless Internet accessibility and for the diversity of its student body. Houston is the nation's fourth most populous city and second lowest in terms of cost of living among major American cities. The city is home to 23 Fortune 500 companies and many other world-class institutions in the areas of energy, medicine, aerospace, manufacturing, and business services. The University's proximity to downtown enables excellent access to these and the many other resources available in Houston and the surrounding region. AA/EOE.
- 122 UNIVERSITY OF MINNESOTA—DULUTH, Department of Accounting, Labovitz School of Business and Economics has a full-time, nine-month, tenure-track, Associate/Assistant/Instructor position available beginning Fall 2011. Job duties and responsibilities include teaching, research, and service. The qualified individual will develop and teach accounting courses in auditing, as well as either financial or managerial accounting areas. Essential qualifications include ABD (with a clear plan for completion by September 1, 2012) from a doctoral program in Accounting or Business Administration with a concentration in Accounting, from an AACSB International-accredited school or major internationally recognized university. For a complete position description and information on how to apply online, visit <http://employment.umn.edu/>; search for Job Requisition 169070. Complete applications will be reviewed beginning February 15, 2011, and continue until the position is filled. The University of Minnesota is an Equal Opportunity Educator and Employer.

- 123 THE UNIVERSITY OF TEXAS AT ARLINGTON, Department of Accounting invites applications for an Endowed Chair (Full Professor) or Endowed Professorship (Associate or Full Professor) position to begin Fall 2011. Job duties include mentoring doctoral students and junior faculty, leading and conducting high-impact accounting research, teaching undergraduate and graduate students, and providing academic and professional service. Candidates should hold a Ph.D. in Accounting, have an outstanding record of scholarly research and publications, an established record of teaching excellence, and a promise of future contributions of an exceptional nature. Professional certification and accounting experience are desirable. Compensation is competitive according to qualifications and experience. Review of applications will begin immediately and continue until the position is filled. This is a security-sensitive position, and a criminal background check will be conducted on finalists. Interested persons should send a current cv, the names and addresses of at least three academic references, demonstrated evidence of excellence in research and teaching, and representative publications or working paper to: Dr. Jennifer Ho, Search Chair Committee, Department of Accounting, College of Business, Box 19468, The University of Texas at Arlington, Arlington, TX 76019. The University of Texas at Arlington does not discriminate on the basis of race, color, national origin, sex, religion, age, disability, veteran status, or sexual orientation in employment or in the provision of services. Effective August 1, 2011, the use of tobacco products (including cigarettes, cigars, pipes, smokeless tobacco, and other tobacco products) by students, faculty, staff, and visitors is prohibited on all UT Arlington properties.
- 124 WASHINGTON UNIVERSITY IN ST. LOUIS, Olin Business School invites applicants to fill faculty positions in the area of Accounting beginning in Fall 2011 to teach courses in the school's Bachelor's, Master's, and /or Ph.D. programs, conduct research, publish in peer-reviewed journals and participate in faculty service activities. Applicants at the Assistant Professor level should have high research potential and have earned their Ph.D. or be close to completion. Seasoned candidates must have a demonstrated outstanding research record and be highly visible in the profession. Applications are being accepted through the Washington University in St. Louis Faculty Recruitment link. Please visit our website: <https://jobs.wustl.edu> and review job # 21014 for further information and to apply. Reviews will begin immediately and will continue until the position is filled. (Hiring is contingent on available funding.) Olin Business School fosters and appreciates ethnic and cultural diversity among its faculty, students, and staff. Applications from men, women, ethnic minorities, veterans, dual-career couples, and individuals with disabilities are encouraged.

For focused results in accounting and tax research, start here.

While the world of accounting information continuously transitions, professionals, educators, and a growing number of accounting students look to reliable resources for researching accounting trends, rules, issues, and history pertaining to the future of accounting.

ProQuest Accounting & Tax, our premier accounting research database, makes research time more effective with centralized access to top accounting journals, reference reports, proceedings, dissertations, and more.

Researchers can instantly cross-search over 2,300 publications, with hundreds in ongoing full-text, plus dissertations and working papers covering accounting and tax topics.

ABI/INFORM® subscribers can enhance their business database library with ProQuest Accounting & Tax, which provides a vast array of additional full-text accounting journals, trade publications, dissertations, and working papers not included in ABI/INFORM.

ProQuest Accounting & Tax includes:

- Highly ranked global and scholarly journals, including all journals from the American Accounting Association
- Accounting, tax management, and auditing trade publications from top publishers such as CCH, Aspen, Thomson Professional and Regulatory Services, Source Media, Tax Management Inc., and Euromoney
- AICPA publications
- International Federation of Accountants (IFAC) standards
- Tax law reviews from top institutions such as New York University, Stanford, Vanderbilt, and University of Virginia
- Tax law and financial law publications from noted sources such as the American Bar Association and American Law Institute
- Global accounting and tax journals from the UK, Canada, Australia, France, Spain, Ireland, and more
- Financial advising, governance, and risk management publications from Incisive Media, CFA Institute, and Association for Financial Counseling and Planning
- Accounting history from publications such as *The Accounting Historian's Journal*
- Over 1,500 accounting dissertations and more than 7,800 working papers from SSRN, plus select conference proceedings



ProQuest Accounting & Tax

Visit www.proquest.com/go/accounting today to learn more, and sign up for a free trial!



CORNERSTONE RESEARCH

Who We Are

Cornerstone Research is a leading economics and financial consulting firm with more than 400 full-time staff members across six offices. Together with an extensive network of nationally prominent faculty who testify as expert witnesses, our staff analyze complex economic, financial, accounting, and marketing issues that arise in litigation. Our culture of growth and collegiality provides excellent career prospects to those who have pursued doctoral studies in accounting, economics, finance, or marketing.

Cornerstone Research is involved in a broad variety of projects, including high-profile legal cases and disputes. Our recent cases have involved a wide variety of accounting and auditing issues, including those related to subprime mortgages, securitizations, fair value, derivative and hedge accounting, revenue recognition, and SOX and other regulatory compliance. Our cases have included allegations against accounting firms, investment banks, corporations, directors, and officers. We have experience in such diverse industries as financial institutions, pharmaceuticals, healthcare, energy, telecommunications, high technology, traditional manufacturing, and retailing. More detail can be found on our Web site at www.cornerstone.com.

Associates at Cornerstone Research

Cornerstone Research provides an interesting and rewarding work environment. Our staff works directly with esteemed faculty experts in a distinctive partnership that combines the strengths of the academic and business worlds. Cornerstone Research provides opportunities for associates to develop as testifying experts if they are interested in such a career path.

Associates are central to casework at Cornerstone Research and are involved in all phases of a project. In the initial stages of a case, associates actively participate in the formulation of the work plan and the analysis of issues. As the case progresses, associates handle the complex aspects of a case directly, while managing and advising analysts. They work with senior staff, experts, and clients to develop case strategy and to determine how best to communicate our findings. In addition, associates actively participate in shaping and implementing the firm's recruiting, training, and practice development strategies. Overall, Cornerstone offers an excellent package of interesting work; exposure to a broad set of issues in finance, economics, and accounting; and a high degree of personal responsibility.

Candidate Profile

We seek candidates who have pursued doctoral studies in accounting with the ability to apply academic research to real-world issues and present concise explanations of complex analyses. The ideal candidate will possess a strong empirical background in financial accounting, cost accounting, or auditing; technical expertise in accounting; and excellent interpersonal skills. CPAs are strongly encouraged to apply.

Interested candidates should submit a cover letter (that includes a statement describing your interest in economics and financial consulting and your location preferences), resume, and a sample research paper to <http://associatecareer.cornerstone.com>. In addition, your application will only be considered complete if it includes three letters of recommendation; please have those sent to us using the information below. Please be sure to rank our offices in order of your location preference, if any. We have offices in Boston, Los Angeles, Menlo Park (CA), New York, San Francisco, and Washington, DC.

Contact

Recruiting Administrator – TAR
Cornerstone Research
599 Lexington Avenue, 43rd Floor
New York, NY 10022

Online Application URL: <http://associatecareer.cornerstone.com>

E-mail: associate-recruiting@cornerstone.com

Cornerstone Research is an Equal Opportunity Employer

Boston, MA • Los Angeles, CA • Menlo Park, CA • New York, NY • San Francisco, CA • Washington, DC

THE ACCOUNTING REVIEW
ISSN 0001-4826

POSTMASTER
Send Address Changes To:

AMERICAN ACCOUNTING ASSOCIATION
5717 BESSIE DRIVE
SARASOTA, FL 34233

PERIODICALS POSTAGE PAID
SARASOTA FLORIDA
AND ADDITIONAL MAILING OFFIC

